INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Theory and Practice of Cross Traffic Estimation via Probes

Sridhar Machiraju — Darryl Veitch — François Baccelli — Jean Bolot

N° 5763

Novembre 2005

Thème COM

ISSN 0249-6399 ISRN INRIA/RR--5763--FR+ENG

ROCQUENCOURT

Theory and Practice of Cross Traffic Estimation via Probes

Sridhar Machiraju [*] , Darryl Veitch [†] , François Baccelli [‡] , Jean Bolot [§]

Thème COM —Systèmes communicants
Projet TREC

Rapport de recherche n° 5763 —Novembre 2005 —25 pages

Abstract: Active probing began by measuring end-to-end path metrics, such as delay and loss, in a direct measurement process which did not require inference of internal network parameters. The field has since progressed to measuring network metrics, from link capacities to available bandwidth and cross traffic itself, which reach deeper and deeper into the network and require increasingly complex inversion methodologies. However, although active probing heuristics are based on queuing systems, to the best of our knowledge, a rigorous probabilistic treatment of probing methods has been lacking. As a result, important issues of system identifiability have been neglected: it is not known, even in principle, what can and cannot be measured in general, nor the true limitations of existing methods. We provide a probabilistic treatment for the measurement of cross traffic in the 1-hop case. We first derive inversion formulae for the law of cross traffic and related processes, and explain their fundamental limits, using an intuitive geometric framework. We then use the resulting insight to design practical estimators for cross traffic, which we test in simulation and validate by using router traces. The estimators perform well, but have natural limitations, which are explained in detail.

Key-words: Measurement, active probing, Internet traffic, router, estimator, identification, queue, network.

[*] Sprint ATL, Burlingame, California, Sridhar.X.Machiraju@sprint.com

[†] ARC Special Research Center for Ultra-Broadband Information Networks (CUBIN), an affi liated program of National ICT Australia, Department of Electrical and Electronic Engineering, The University of Melbourne, Australia, d.veitch@ee.unimelb.edu.au

[‡] INRIA-ENS, 45 rue d'Ulm 75005, Paris, France. francois.baccelli@ens.fr

[§] Sprint ATL, Burlingame California, Jean.C.Bolot@mail.com

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)
Téléphone : +33 1 39 63 55 11 —Télécopie : +33 1 39 63 53 30

Théorie et Pratique de l'Estimation du Trafic Transversal par des Sondes

Résumé : Les méthodes de sondes actives ont initialement été introduites pour l'évaluation de propriétés vues de bout en bout par les flots, telles que les délais ou les pertes. Ces méthodes ont ensuite été étendues à l'estimation de paramètres internes du réseau, tels que la capacité des liens et plus récemment la bande passante disponible ou encore les propriétés statistiques du trafic transversal. Cette vision de plus en plus détaillée de l'intérieur du réseau traversé nécessite des méthodes d'inversion de complexité croissante. Même si les heuristiques à base de sondes actives sont souvent fondées sur des méthodes des files d'attente, il n'existe pas encore à notre connaissance de cadre théorique permettant de poser les questions d'identifiabilité, et on sait donc très peu actuellement sur ce qui peut être effectivement être estimé dans ce cadre. Dans cet article, nous proposons un cadre probabiliste qui permet de poser les questions d'identification de la loi du trafic transversal, dans le cas particulier d'un seul saut. Nous établissons une formule d'inversion permettant d'analyser les propriétés du trafic transversal et nous montrons les limitations de cette méthode d'inversion au moyen d'arguments géométriques simples. Nous utilisons cette formule d'inversion pour définir des estimateurs de la loi du trafic transversal que nous testons par simulation et sur des traces de trafic obtenues par des mesures sur des routeurs. Ces estimateurs sont utilisables dans la pratique, avec des limitations que nous expliquons en détail dans l'article.

Mots-clés : Mesure, sonde active, trafic Internet, routeur, estimateur, identification, file d'attente, réseau.

1 Introduction

Active probing has become one of the main ways in which the performance of IP networks such as the Internet are measured. In active probing, a stream of *probe packets* are injected from a source at the edge of the network, and collected at cooperating receiver(s). From the known packet sizes, and the measured timestamps of emission and reception, information on both static network parameters and traffic conditions can be inferred.

The literature initially focused on link bandwidth estimation, [9, 3, 16], and in particular that of the smallest link, the *bottleneck bandwidth*. More recently, estimation of *available bandwidth* along a path, which is a function not only of the physical network but also of all the *cross traffic* impacting on the probes, has attracted considerable interest [10, 19]. Far less attention has been paid to the idea of using probing to measure queueing processes in routers, or properties of the cross traffic itself (see however [20, 18, 15]), although there is a general acceptance that this can be a very difficult problem over multiple hops. In fact, there are deeper underlying issues of system identifiability which equire investigation. There is currently no consensus on what can and cannot in fact be measured. For example, there are no results available capable of determining what the true limitations of existing techniques actually are, or if their goals are even feasible. One way of thinking about the problem is in terms of ambiguity, are there different cross traffics which can give rise to the same observations? We will show that the answer is yes, however the more important question is whether this is true not just of sample paths but of the underlying distributions which define the cross traffic.

In this paper we make what we believe to be one of one first steps toward the understanding of the in-principle potential and limitations of active probing methods. We do this by looking in some detail into the problem of measuring the distribution of the cross traffic process through the histories probes accumulate in traversing a hop. Since the probes interact with the cross traffic via queueing systems, it can be viewed as an 'inverse queueing problem'. It necessitates considering, in a joint fashion, the statistics of both the queueing process and that of the arriving traffic, and is very different from the traditional questions studied in queueing theory. We show that, in a simple yet well motivated 1-hop setting, complete inversion may be possible, or impossible, depending on details of the traffic itself, and we define what this means precisely. Previous network inference work which is statistically rigorous, notably the network tomographic literature (see [22] and references therein), are based on selected abstractions of probe delay behaviour. They do not deal with a true queueing system model as we do here.

The second aim of the paper is to use the insights gained through the in-principle inversion procedure we develop, to define statistical estimators for the distribution of the cross traffic which can be used in practice. To our knowledge this is one of the first works where a rigorous probabilistic treatment has been given to such a problem. We provide a number of different estimators, suitable for different circumstances, and explain in detail the principles on which they work, and when and why they may fail. We make extensive use of geometric arguments and illustrations to give an intuitive account. We evaluate their performance in Monte Carlo simulations using cross traffic processes which are reasonable first models of Internet traffic. We also make use of a detailed Internet traffic trace, to give an indication of the utility of the method under realistic conditions which deviate from the technical assumptions, described below, used in the inversion. In addition, we perform simultaneous probing experiments and accurate passive capture in the Internet backbone, and use it to evaluate the performance of the estimators under real world conditions.

Although in general a single hop is of limited usefulness as a model of a multi-hop route, we explain how our approach has advantages over existing methods based on this idea which widen its applicability.

We use the prevailing hop model consisting of a FIFO queue to which both cross traffic packets and probes effectively arrive instantaneously, but flow out deterministically as they are serialised onto the output link. In this picture, the service time of the traffic is associated to the input process itself, rather than the server, and arrival and departure times are measured from the end of packets. This abstraction of hop behaviour is appropriate in today's Internet where store and forward router architectures are common, with fast switch fabrics where through-router delays are concentrated in output buffers. It was recently validated using real data collected at the input and output interfaces of a router in the Sprint backbone network [5].

2 In-Principle Inversion

We consider a simplified problem consisting of a single hop, whose FIFO single server queue is characterised by a deterministic service rate μ, and an infinite buffer. We take the probes to be of constant size chosen to be p bytes, corresponding to $x = p/\mu > 0$ seconds of workload. We will comment on the role of x as a parameter.

Let $\{T_n\}$ and $\{T'_n\}$ be the sequence of arrival and departure times respectively of the probes to the queue. Since the $\{T_n\}$ can be chosen by the prober and are assumed known, the raw data of a probing experiment are the departure times, or equivalently, the end-to-end hop *delays* $\{D_n = T'_n - T_n\}$. The broad problem can now be stated as follows:

Given a knowledge of the measured delays, what can be learnt about the probability laws governing the cross traffic?

It is convenient to describe the input traffic in terms of a *random measure* A, whereby the workload (measured in seconds after dividing by μ) arriving to the queue in a time interval I is denoted by the random variable $A(I)$. In this way we include point or continuous arrivals in a unified and general framework.

Our aim is to recover as much information as possible about the cross traffic described by A. For this to be feasible, the statistics of the system should not change fundamentally over time, and the probes must be able to collect representative samples of them. The corresponding technical assumptions are that A is stationary (i.e. for all intervals I, $A(t+I)$ has a law that does not depend on t) and that the sequence of end-to-end probe delays is stationary and ergodic.

For the mathematical development in this section and the next, it is necessary to state specific joint assumptions on A and the

$\{T_n\}$. For the purposes of this paper we will assume one of the following two models. We show in the appendix that the stationarity and ergodicity assumptions are satisfied in each.

- Model 1: $\{T_n\}$ is an arbitrary renewal point process. It is independent of A, which has stationary independent increments, and infinite support for all t;
- Model 2: $\{T_n\}$ is deterministic (i.e. periodic) with fixed interarrival time t. The sequence of stochastic processes $\{A([it, it+v)), 0 \le v \le t\}_{i \in \mathbf{N}}$ is independent and identically distributed. The following technical assumption is also made: $A[0,t) = 0$ with a positive probability and achieves values larger than $t-x$ with a positive probability.

The raison d'être of these two models and their differences and relative merits will be discussed later and in particular in Section 3.

Throughout this section we will for simplicity use Model 1, which includes as an important special case cross traffic packets arriving as a Poisson process, where the packet sizes are independently drawn from some distribution. In fact our approach is general enough to be applied in other situations, for example probe streams do not have to be renewal; one could for instance consider the interarrivals $\{T_{n+1} - T_n\}$ to be Markov.

It is important to understand that our approach takes the invasive effect of probes fully into account. In no way is a low rate probing stream required (either locally or on average) by the inversion methods presented, or the estimators based on them.

We begin by giving the generic equations governing the system. We then derive recursion relations connecting the observed delays, and go on to explain the principle of the inversion approach under our statistical assumptions of Models 1 and 2. The relationship between the results here and questions of identifiability will be left to the next section.

2.1 An Approach to Inversion

Using standard queueing theory for FIFO queues (for example see [2]), the equation describing probe delays can be written as

$$T'_{n+1} = x + \big[(T'_n + A[T_n, T_{n+1})) \vee \sup_{v \in [T_n, T_{n+1}]} (v + A([v, T_{n+1}))\big], \quad (1)$$

where $x \vee y$ denotes the maximum of x and y. The left hand argument of $\vee$ dominates when the probes are in the same busy period, as then the departure time of probe $n+1$ is simply determined by the cross traffic $A([T_n, T_{n+1}))$ filling the space between them. If this is not the case then the supremum, which allows the calculation of the waiting time experienced by probe $n+1$, will dominate instead.

Subtracting T_{n+1} from both sides of the equation above, a recursive relationship emerges for the delay $D_n = T'_n - T_n$. Instead of using absolute delay however, it is convenient to work with the time series $\{R_n\}$, where $R_n = D_n - x \ge 0$ is the excess delay above the minimum value of x, the probe service time. In terms of the R_n the recursion becomes

$$R_{n+1} = (x + R_n + C_n) \vee B_n \,, \quad (2)$$

where

$$C_n = A([T_n, T_{n+1})) - (T_{n+1} - T_n), \quad (3)$$

$$B_n = \sup_{v \in [T_n, T_{n+1}]} (A([v, T_{n+1})) - (T_{n+1} - v)) \,. \quad (4)$$

Note that B_n and C_n are functionals of the cross traffic over the interval $[T_n, T_{n+1})$ *only*, and are neither influenced by the probes nor by the queue state. The following important relationships hold:

1. $B_n \ge 0$ (take $v = T_{n+1}$),
2. $B_n \ge C_n$ (take $v = T_n$),
3. $B_n \le C_n + (T_{n+1} - T_n)$ (since $B_n \le A([T_n, T_{n+1}))$).

We can interpret C_n as the net work that arrives in $[T_n, T_{n+1})$, and it takes values in $(-(T_{n+1} - T_n), \infty)$. Thus, C_n gives information on the *integral* of the cross traffic over a typical probe inter-arrival, whereas B_n gives some information on the *peak*. More precisely, B_n is the system workload that would be seen at time T_{n+1} if we considered *only* the cross-traffic arriving in $[T_n, T_{n+1})$. For example, $B_n - C_n$ is maximized when the traffic arriving over $[T_n, T_{n+1})$ occurs in a burst just before T_{n+1}. Another example from network calculus is given in the appendix.

The technical conditions listed as Model 1 or Model 2 have the following important consequences:

1 R_n, which is determined by what the n-th probe encounters when it arrives in the queue at time T_n, is independent of $(B_n, C_n, (T_{n+1} - T_n))$, which is a function of the traffic arriving after T_n;

2 The sequence $\{R_n\}$ is an ergodic Markov chain, and therefore admits a (unique) stationary and ergodic regime under natural rate conditions (see appendix).

In what follows, we will assume the system to be in its stationary regime.

We now describe the core ideas of the inversion procedure. Since we know the joint process of the excess delay $\{R_n\}$ and the probe arrivals $\{T_n\}$, we know in particular the conditional law of R_{n+1} given $T_{n+1} - T_n$ and R_n. Rephrased in terms of sample paths, we have a complete history of the random variables $\{R_n\}$ and $\{T_n\}$, and can therefore pick out those n which correspond to $T_{n+1} - T_n = t$ and $R_n = r$, for any fixed $t > 0$, $r \ge 0$, and consider the corresponding R_{n+1} values. The conditions on ergodicity given above ensure that all values of r are visited arbitrarily often (for a large enough number of probes), and since in addition (B_n, C_n) is independent of R_n, all possible 3-tuples (R_n, B_n, C_n) will occur arbitrarily often. The conditional delays correspond to a sequence of (B, C) pairs sharing the same fixed t, of which a representative element can be written

$$C = A([0,t)) - t, \quad (5)$$

$$B = \sup_{0 \le v \le t} (A([v,t)) - (t - v)). \quad (6)$$

The corresponding recursion relation linking the residual delay R of such a probe to the residual delay S of the next probe, which arrives a time t later, is given by

$$S = [x + R + C] \vee B. \quad (7)$$

We now again use the property that R is independent of (B, C). It follows that S is determined by the observable R, and the independent *unobservable* joint distribution of (B, C). As the unknown is now just a 2-dimensional distribution, this is an important increase in tractability brought about by fixing t and by the technical assumptions. If we can determine this distribution, then we know in particular the marginal distribution of C, and thereby the distribution of $A(t) = C + t$. Here we have written $A(t)$ as a shorthand for $A([0, t))$.

The above was for a single t fixed. For this sub-problem the task has been reduced to recovering the unobservable joint distribution of (B, C), based on the observable (and therefore, in this section, known) joint distribution of (R, S). To proceed with the inversion, we henceforth assume that all variables, including time, are discrete. This assumption is not essential, as the discretisation can be made as fine as we wish, and in applications using real data, discretisation is in any case unavoidable.

We denote the discrete density and the 2-dimensional cumulative distribution function (CDF) of (B, C) respectively by

$$h(k, l) = P(B = k, C = l), \tag{8}$$
$$H(k, l) = P(B \leq k, C \leq l), \tag{9}$$

in the domain of definition $k \geq 0, l \geq -t$.

We write $c(l)$ and $C(l)$, $l \geq -t$, for the density and CDF respectively of the marginal corresponding the variable C, and similarly $b(k)$ and $B(k)$, $k \geq 0$, for B. In addition, it is convenient to define the following 'approximations' to $c(l)$ and $b(l)$:

$$c(k, l) = \sum_{i=0}^{k} h(i, l) \tag{10}$$
$$b(k, l) = \sum_{i=-t}^{l} h(k, i). \tag{11}$$

The sets of k, l pairs in the sums defining $c(\cdot, \cdot)$ and $b(\cdot, \cdot)$ appear in Figure 1 as finite horizontal and vertical bars respectively.

The relationships 1–3 listed above imply that $h(k, l) = 0$ outside of the diagonally oriented 'feasible strip' defined by

$$\text{feasible strip:} \quad (k, l) : k - t \leq l \leq k, \quad k \geq 0, \tag{12}$$

as seen in Figure 1. This implies in particular the important fact that $c(k, l) = c(l)$ as soon as $k \geq l + t$.

2.2 Inversion Expressions

Let $f_r(s) = P(S = s | R = r)$. We can write down this conditional probability by accounting the cases when either the left or the right hand arguments in Equation (7) equals s. Due to the independence of R from (B, C), this can be simply written as

$$\begin{aligned} f_r(s) &= P(B \leq s, C = s - r - x) \\ &\quad + P(B = s, C \leq s - r - x - 1) && (13) \\ &= c(s, s - r - x) + b(s, s - r - x - 1) && (14) \\ &= H(s, s - r - x) - H(s - 1, s - r - x - 1). && (15) \end{aligned}$$

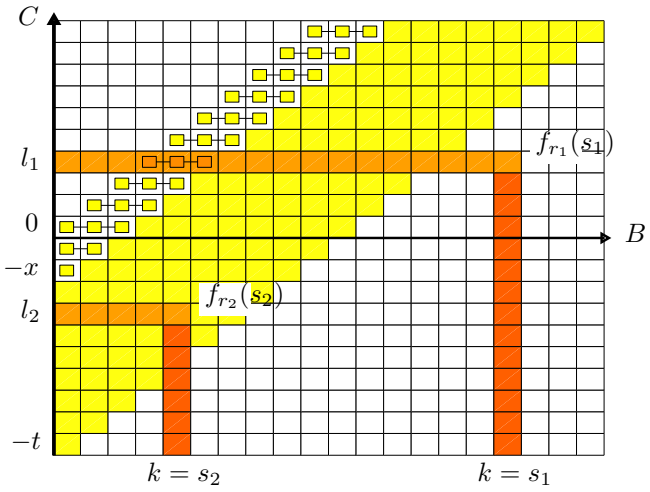

Figure 1: The domain $\{k, l\}$ where the joint density $h(k, l)$ of (B, C) vanishes is shown as white. The support of (B, C) is the strip $k - t \leq l \leq k$, $k \geq 0$ shown as the light coloured band. An observation of $(R, S) = (r, s)$ corresponds to a $(B, C) = (k, l)$ value lying inside an angle shaped set with corner at $(k^*, l^*) = (s, s - r - x)$. Two angle sets are shown (shaded), corresponding to Class 1 (corner outside the strip, $k = s_1$), and Class 2 (corner inside, $k = s_2$). The region where aggregates of $x + 1 = 3$ atomic masses are connected is the *exclusion zone* where individual h values cannot be directly determined.

This probability corresponds to a sum of $h(k, l)$ over an 'angle shaped' set, with corner at

$$(k^*, l^*) = (s, s - r - x).$$

Two examples of angle sets are illustrated in Figure 1. A particular observation of $S = s$ given $R = r$ corresponds to a $(B, C) = (k, l)$ value, which, although unobservable, must lie inside the angle set defined by (r, s). We see that the available information concerning $h(k, l)$ comes in the form of the probabilities, given by $f_r(s)$ for different observed (r, s), of falling into different angle sets. All other knowledge of the joint density must be obtained by combining such sets in different combinations. We will now describe how this works in more detail with the aid of Figure 1.

A given (k, l) value may be included in many angle sets corresponding to different (r, s), however the mapping between (r, s) and the corner (k^*, l^*) is linear, and hence uniquely invertible: $(r, s) = (k^* - l^* - x, k^*)$. Consider then the possible locations of the 'corners'. For a fixed r, as s is increased the corresponding corner values $(k^*, l^*) = (s, s - r - x)$ move upward, tracing out a line parallel to the main diagonal. As r decreases these diagonals translate upward, however the highest of these, corresponding to $r = 0$, is not the upper boundary of the strip, but lies below it on the line $l = k - x$. We discuss the consequences of this for invertibility in the next section.

We now derive inversion expressions, in three classes. The first class aims to directly determine $c(l)$. It is simple and intuitive, being based on restricting to cases where the queue is known to be 'linear', that is when it is *certain* that the probes

share the same busy period. The second is based on the idea that conditioning on linearity may be too strong, and require too much data in practice, motivating expressions for the approximation $c(k,l)$ to $c(l)$, justified by the idea that we can ignore zones where the density is likely to be small. The third class is totally different, being based on a transform viewpoint rather than operating purely in the time domain. However, it relies fundamentally on the prior time domain results to obtain information on A for different values of t. Its advantage is that it allows us to combine different t values in a natural way. It is however based firmly on Model 1. For example it does not hold for Model 2.

2.2.1 Inversion: Class 1

The first inversion expression uses the observation from section 2.1 that $B \leq C + t$, which implies that $B \leq x + r + C$ if $R = r \geq t - x$. Hence, for $r \geq t - x$, Equation (7) implies that

$$\begin{aligned} P(S = s|R = r) &= P(x + R + C = s|R = r) \\ &= P(C = s - r - x|R = r) \\ &= P(C = s - r - x), \end{aligned} \tag{16}$$

which is a function of the *delay variation* $u = s - r$. The last step follows from the independence of R and (B, C) established above. Thus, for each fixed $R = r$ obeying $r \geq t - x$ we have

$$c(l) = f_r(l + r + x). \tag{17}$$

In terms of angle sets, The above simply corresponds to taking a corner (k^*, l^*) with $l^* = s - r - x$, and k^* large enough so that the horizontal component of the angle completely traverses the strip (see the rightmost angle in Figure 1). The vertical component in such cases falls below the strip, and thereby contains zero probability.

Since the above expression is true for many different r values, it is desirable to combine them, as this would make better use of data in the practical case. A general linear combination can be written as

$$c(l) = \sum_{r=t-x}^{\infty} a_{r-(t-x)} f_r(l + r + x), \tag{18}$$

where the a_i are any set of non-negative weights that sum to unity. Intuitively, a good choice is to select weights that reflects the data available: $a_{r-(t-x)} = P(R = r|R \geq t - x)$. It turns out that the resulting expression is the same as if we had set out, looking across different r values, to explicitly collect together all relevant observations with constant u:

$$\begin{aligned} &P(S - R = u|R \geq t - x) = P(x + C = u|R \geq t - x) \\ &= P(C = u - x) \\ &= \sum_{r=t-x}^{\infty} P(S - R = u, R = r)/P(R \geq t - x) \\ &= \sum_{r=t-x}^{\infty} P(S - R = u|R = r)P(R = r))/P(R \geq t - x) \end{aligned}$$

which is of the form of Equation (18). Defining $g_r(u)$ to be $P(S - R = u|R \geq r)$, we get:

$$c(l) = g_{t-x}(l + x). \tag{19}$$

The collection of (r, s) values used in this expression is illustrated in Figure 2(a), where the shading indicates the corresponding weight.

The above expressions essentially choose observations corresponding to large delay values of the first probe in a pair. The large delay ensures that the queue is busy until the next probe arrives. The difference between such a delay and the next delay is used to estimate the distribution of cross-traffic arriving between them.

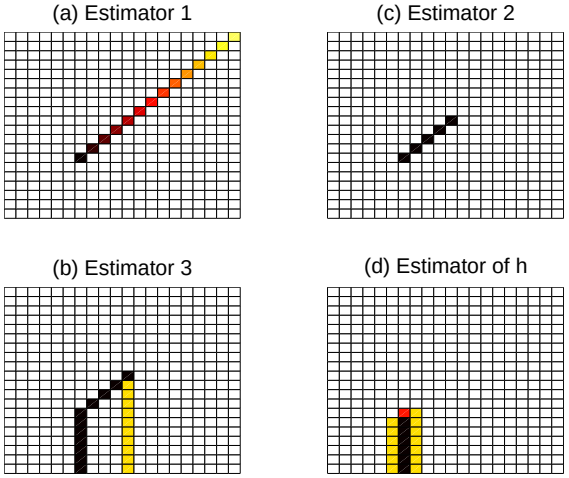

Figure 2: The various inversion expressions use different portions of the (R, S) space, shown here. The corresponding equation numbers are: (a) $\hat{c}_1$: (19) (shading indicates weights used); (b) $\hat{c}_2$: (22) (correction terms give the vertical components); (c) $\hat{c}_3$: (24) (uniform weighting over N values of r); (d) $\hat{h}$: (33) (shading indicates term type).

We conclude with a comment on the role of x. Since increasing x widens the exclusion zone, and more generally increases delays *without impacting* on the density $h(k,l)$, it serves to increase the available range of r values, thereby improving the applicability of expressions in Class 1.

2.2.2 Inversion: Class 2

The expressions for $c(l)$ above relied on $r \geq t - x$. Thinking ahead to practical situations with limited data, such values may be rare or even entirely unavailable. However, if $h(k,l)$ is sufficiently concentrated at small k and l, then we may be able to use $c(k,l)$, defined in Equation (10), as a valid surrogate for $c(l)$. We accordingly consider expressions based on $c(k,l)$.

Let

$$F_r(s) = \sum_{i \leq s} f_r(i) = P(S \leq s|R = r)$$

be the CDF corresponding to $f_r(\cdot)$, which is observable. Using Equation (15), we can solve for $c(k,l)$ as follows:

$$\begin{aligned} c(k,l) &= H(k,l) - H(k,l-1) \\ &= F_{k-l-x}(k) - F_{k-l-x+1}(k) \\ &= \sum_{i=0}^{k} [f_{k-l-x}(i) - f_{k-l-x+1}(i)] \\ &= f_r(r+l+x) + F_r(r+l+x-1) \\ &\qquad - F_{r+1}(r+l+x), \end{aligned} \tag{20}$$

where $r = k-l-x$. The second and third lines follow immediately from the fact that $H(k,l) = F_{k-l-x}(k)$ is by definition a 'rectangle' set that can be decomposed into nested angle sets.

Comparing Equation (17) to (20), we see that the first term is still $f_r(r+l+x)$, however it is no longer equal to $c(l)$ unless $r \geq t-x$, in which case the extra terms in Equation (20) are equal and cancel.

Equation (20) provides us with an exact expression for $c(k,l)$ which we can think of as an approximation to $c(l)$. We continue by considering different assumptions on which components of Equation (20) are considered to be negligible, resulting in new expressions.

Weak Assumption:
Consider Equation (20) for a given l. If in fact $h(k,l)$ is negligible for $k > k_l$, where $k_l \in [l, l+t-1]$ lies inside the strip, then it is as if the strip were narrower, and $c(l) = c(k,l)$ whenever $k \geq k_l$. We refer to this as the *weak assumption*. It is reasonable to expect that it holds in many cases, at least for sufficiently large l, as increasing k corresponds in some sense to 'tail' events of low probability. It leads to the following approximation for each fixed r obeying $r \geq r_l = k_l - l - x$:

$$\begin{aligned} c(l) &\approx F_r(r+l+x) - F_{r+1}(r+l+x), \\ &\quad \text{assuming } h(l+r'+x, l) = 0 \quad \forall\, r' > r. \end{aligned} \tag{21}$$

Equation (21) uses only a single $r \geq r_l$. As in the previous section, it makes sense to combine different values to reduce variabililty. Using uniform weights results in a satisfying cancellation of $f_r(s)$ terms, which would not occur if weighted averaging were used. Using r now as a parameter obeying $r \geq r_l$, we obtain:

$$\begin{aligned} c(l) &= \frac{1}{N} \sum_{r'=r}^{r+N-1} [F_{r'}(r'+l+x) - F_{r'+1}(r'+l+x)]. \\ &= \frac{1}{N} \sum_{r'=r}^{r+N-1} f_{r'}(r'+l+x) + \\ &\quad \frac{F_r(r+l+x-1) - F_{r+N}(r+N+l+x-1)}{N} \end{aligned} \tag{22}$$

The first term is the average of N expressions of the form of Equation (20), whereas the second includes $h(k,l)$ values that are not affected by the weak assumption. In Figure 2(b) the (r,s) values required by the extra terms appear as the vertical lines.

Strong Assumption:
If k_l is such that $h(k_l, l)$ can be considered a tail probability, it is reasonable to expect that this may also be true of $h(k_l, l')$ for l' values close to l. This motivates the following *strong assumption*, which in addition to the weak assumption, supposes that $h(k', l')$ vanishes for all $l' < l$ when $k' \geq k_l$, or equivalently $r' \geq r_l$. In other words, all elements directly below, and below and to the right of, the point (k_l, l). It is not difficult to see that, for a fixed r obeying $r \geq r_l = k_l - l - x$, this leads to

$$\begin{aligned} c(l) &\approx f_r(r+l+x), \\ &\quad \text{assuming } h(l+r'+x, l) = 0 \quad \forall r' > r \\ &\quad \text{and } h(l'+r'+x, l') = 0 \quad \forall r' \geq r,\ l' < l. \end{aligned} \tag{23}$$

As we did with Equation (21), we can average N consecutive values starting from a given $r \geq r_l$, yielding

$$c(l) \approx \frac{1}{N} \sum_{r'=r}^{r+N-1} f_{r'}(r'+l+x). \tag{24}$$

The leftmost angle in Figure 1 is an example of the terms (angles) in this sum whose corner lies inside the strip. The stronger assumption has resulted in the loss of the difference term of Equation (22). The corresponding plot Figure 2(c) shows that the vertical lines have vanished, leaving a diagonal set similar to Figure 2(a), only with uniform weights.

We can also perform averaging with a natural set of weights as we did with the class 1 inversion methods. Recall that $g_r(u) = P(S - R = u | R \geq r)$, and let $p_r(r')$ denote the conditional probabilities $P(R = r' | R \geq r)$, interpreted as a set of weights which sum to 1. Using the strong assumption one can show that

$$\begin{aligned} c(l) &\approx \sum_{r'=r}^{\infty} f_{r'}(r'+l+x) p_r(r') \\ &= \sum_{r'=r}^{\infty} P(S = l+x+R | R = r') P(R = r' | R \geq r) \\ &= \sum_{r'=r}^{\infty} P(S - R = l+x | R \geq r) \\ &= g_r(l+x). \end{aligned} \tag{25}$$

This is formally the same expression as Equation (19)! however now r is smaller than $t-x$, and the corners of the angles set involved are inside the strip.

Just as we did for $f_r(s)$, we can define a CDF corresponding to $g_r(u)$, and establish the following identity:

$$G_r(u) = \sum_{i=x-t}^{u} g_r(i) = P(U \le u | R \ge r), \quad (26)$$

$$= \sum_{l'=-t}^{l} g_r(l'+x),$$

$$= \sum_{l'=-t}^{l} \sum_{r'=r}^{\infty} f_{r'}(r'+l'+x) p_r(r'),$$

$$= \sum_{r'=r}^{\infty} p_r(r') \sum_{l'=-t}^{l} f_{r'}(r'+l'+x),$$

$$= \sum_{r'=r}^{\infty} p_r(r') F_{r'}(r'+l'+x), \quad (27)$$

$$= \sum_{r'=r}^{\infty} p_r(r') H(r'+l+x, l), \quad (28)$$

which shows that the CDF corresponding to $g_r(u)$ can be viewed as a weighted sum of rectangle sets, with the same (l independent) weights as before.

For this class, increasing x has the advantage that a fixed (r, s) observation now corresponds to an angle which is lower in the strip (corresponding to larger r values), whereas the density h, being independent of probes, has not changed. Hence, this same (r, s) pair is now more likely to fall into a region where the weak and strong assumptions hold than before. For the same reason, higher r values now have higher probability, allowing larger r' values to be used for a fixed number of observations. Each of these advantages is consistent with the intuition that increased local invasiveness increases the chances of probes being in the same busy period, where information can be extracted about $c(l)$. The disadvantage is the the widening of the ambiguity zone of course decreases the observability of h.

2.2.3 Inversion: Class 3

In this section we assume Model 1, that is that A has independent increments and the probes arrive as a renewal process, and we return to the continuous time and space framework. The observation at the heart of this section is that such a process is uniquely characterized by its Lévy exponent $\alpha(u)$. One way of accessing $\alpha(u)$ (provided natural regularity conditions are satisfied) is through the Laplace transform of the Lévy process, defined by $\phi_u(t) = E[e^{-uA(t)}]$. Independent increments implies that the transform obeys

$$\phi_u(t+s) = E[e^{-uA(t+s)}] = E[e^{-uA(t)} e^{-uA(s)}] = \phi_u(t)\phi_u(s), \quad (29)$$

the canonical solution to which is $\phi_u(t) = e^{-\alpha(u)t}$. For a Poisson process with intensity λ for example, $\alpha(u) = \lambda(1 - e^{-u})$. More generally, in the case of the simple model of Internet traffic discussed in Section 2, namely when random sized packets arrive as a Poisson process, then

$$\alpha(u) = \lambda(1 - F(u)),$$

where $F(u)$ is the Laplace transform of the service time distribution. We can therefore determine $\alpha(u)$ simply by inverting the exponential solution:

$$\alpha(u) = \frac{-1}{t} \log E[e^{-uA(t)}]. \quad (30)$$

Through $\alpha(u)$, we obtain full knowledge of the process A simply by knowing the marginal $A(t)$ at a single t value.

Although in principle only a single t is needed, for practical estimation purposes it is pertinent to consider multiple values. For example with renewal probe arrivals, it is natural that many different t values occur. These could be combined linearly using Equation (30) to form a new expression for $\alpha(u)$:

$$\alpha(u) = -\int_0^{\infty} \frac{b(t)}{t} \log E[e^{-uA(t)}]\, dt, \quad (31)$$

where the $b(t)$ is a weighting function such that $\int_0^{\infty} b(t)\, dt = 1$. It is reasonable to choose this function to be equal to $a(t)$, the inter-arrival time density between probes, so that greater weight is given to t values which occur more often. Equation (31) is remarkable as it combines information from different t values in a natural way, something that is very difficult to do using the expressions of Classes 1 and 2. It therefore makes better use of available samples, which will be useful in practice, especially for passive probing.

Although the above is transform based, the marginal $A(t)$ must first be determined, and the only means we have to achieve this is to use the Class 1 and 2 inversions described above. There are therefore many possible expressions for $\alpha(u)$ in terms of observables, corresponding to combining different choices of weights in Equation (31), together with the variety of expressions from Classes 1 and 2.

3 Ambiguity, Invertibility & Identifiability

As already mentioned, our aim is to identify, that is to fully know, the nature of the cross traffic based on the observations of probes. If this cannot be done in a sample path sense, then we wish at least to determine in full the stochastic laws describing the statistics of the traffic. It should be clear from Equation (2) that the best thing that can be identified in the present context is the (B_n, C_n) sequence. In view of the fact that this sequence is i.i.d. under our assumptions, the law of the process will be identified simply if one can determine the 2-dimensional joint distribution or *law* of B and C, conditional on t, from the observations. We shall say that the system is invertible (resp. pathwise invertible) if this law (resp. sequence) can be estimated from the observations and that it is ambiguous (resp. pathwise ambiguous) if more than one such law (resp. sequence) is compatible with the observations.

It is easy to see that our system is pathwise ambiguous, namely that different cross traffics can give rise to the same sequence of observed probe delays, even in a single hop model.

To show pathwise ambiguity, consider a case where two probes share the same busy period, and where a number of cross traffic packets arrive and are 'trapped' between them. The arrival order of the cross traffic packets could be changed, or a number of them could be replaced with a smaller number with an

equivalent total service time, without altering the delays of either probe. Similarly, when two successive probes do not share the same busy period, these probes have actually no way of sensing what happens at *every* point along the interval of time separating them. More precisely, all modifications of cross traffic which leave unaffected the (distinct) busy periods containing the probes in question lead to the same pathwise observations. So, if cross traffic is such that pairs of successive probes are *never* in the same busy period, then there is clearly a wide variety of cross traffic processes that lead to the same pathwise observations.

It is not difficult to see that when probes never belong to the same busy period, then there is ambiguity in law too. We shall see below that even if one excludes this case and assumes that there are infinitely many pairs of successive probes that belong to the same busy period and that moreover probe delays form an ergodic sequence (on the positive half-line), then there is still ambiguity in law for Model 2.

Hence, at least under the assumptions of Model 2, it is not possible in general to identify the statistics of cross traffic via probing, even in the one hop model.

3.1 Ambiguity of Model 2

The results of Sections 2.2.1 and 2.2.2 were based on what could be obtained about $A(t)$ using the distribution of S given R for a given probe separation $T = t$. The most that could be hoped for is a complete recovery of the density $h(k,l)$ of the joint variable (B,C) which underlies $A(t)$. We now consider to what extent this can be achieved.

Since $h(k,l) = c(k,l) - c(k-1,l)$, Equation (20) provides us with a convenient formal way of calculating the joint density:

$$h(k,l) = \sum_{i=0}^{k-1} [2f_{k-l-x}(i) - f_{k-l-x-1}(i) - f_{k-l-x+1}(i)] + [f_{k-l-x}(k) - f_{k-l-x+1}(k)] \quad (32)$$

$$= F_{k-l-x}(k) + F_{k-l-x}(k-1) - F_{k-l-x+1}(k) - F_{k-l-x-1}(k-1). \quad (33)$$

Equation (33) provides $h(k,l)$ using three values of r, namely $k-l-x$ and $k-l-x\pm 1$, and several values of s. The collection of (r,s) values required is illustrated in Figure 2(d).

Since $F_r(s)$ is undefined for $r < 0$, the above expression for $h(k,l)$ can be used only when $k - l - x \geq 1$. Hence, we in fact cannot determine individual $h(k,l)$ values for $k - l \leq x$! We call this region, marked with smaller rectangles in Figure 1, the *exclusion zone*. However, Equation (20) shows that, for fixed l, we do know $c(l+x,l)$ which is the sum of $h(k,l)$ over $l \leq k \leq l+x$, that is the mass in an aggregate traversing the width of the exclusion zone (an exception occurs when $l = -x$, where $h(0,l) = c(0,l)$ *is* known from Equation (20)). Note that other aggregates involving (k,l) values in the zone cannot be calculated. In particular, the marginal $b(k)$ of B cannot be determined. In other words, Model 2 is law ambiguous. The probe size x here plays a key role. Smaller x reduces the exclusion zone and enables h to be determined more fully.

The above can also be explained geometrically in terms of angle sets. The highest placed angles are those corresponding to $r = 0$, whose corners correspond to the diagonal comprising the lower edge of the exclusion zone. Since points in the interior of the zone, that is $l \geq k - x - 1$, cannot be corners, it follows that for a given $l \geq -x$ the *only* angles passing through points in the zone are those whose horizontal members align at $l = s - r - x$. Consequently, it is impossible to resolve the $h(k,l)$ values in the interior of the zone.

To understand *why* the corners cannot lie in the interior of the exclusion zone, note that the horizontal component $c(s, s-r-x)$ of the angle set (r,s) (refer to Equation (15)) corresponds to (k,l) pairs such that the two probes share the same busy period, whereas the vertical component $b(s, s-r-x-1)$ contains scenarios where they do not. Because of the invasive impact of the probe size x however, they **must** be in the same busy period if $l \geq k - x$, which is precisely the definition of the exclusion zone.

3.2 Invertibility of Model 1

We saw in Section 2.2.3 that in the case when the measure A has stationary independent increments, it is possible in principle to determine, from the knowledge of the marginal $A(t) = A[0,t)$ for a single t, the function $\alpha(u)$ known as the Lèvy exponent which fully characterises the system. It follows that, from this single t, we learn not just the marginal $A[0,t)$, but all there is to know about the entire measure A. As a result, we also know in principle the joint law of (B,C), for all t. Thus in this case there is no fundamental barrier to system identifiability. The ambiguities found in Section 2.2.3 do not hold because of the following property of processes with independent increments: the Lévy exponent (which here is identifiable) fully determines the law of the process.

There is no contradiction between the ambiguity present for a fixed t, corresponding to the existence of the exclusion zone described above, and the resolution of that ambiguity in the context of Model 1. The exclusion zone relates to what can be *directly* inferred on the basis of measurements made at a constant t, with no additional information. Recall that the law of C, and therefore of $A(t)$, can indeed be recovered based on such observations. The structure inherent in Model 1 effectively parameterises the problem, so that (B,C) is contrained to be of a certain form such that if the law of C is known, then so is that of (B,C). The same logic does not hold for Model 2, quite simply because no additional structure is given.

3.3 Comments on Model 1 and 2

The processes from Model 1 with independent increments are quite general in that they can include both a continuous component consisting of a diffusion process (e.g. Brownian motion), and a jump component consisting of a compound arrival Poisson point process. The following process is a relevant example of a pure jump component: packets arrive instantaneously at Poisson epochs, bringing with them a random amount of server load correponding to independent samples of some packet size distribution.

In Model 2, no specific assumption is made on the law of A within each of the intervals $[it,(i+1)t)$, and it is therefore more

general than Model 1 in this respect. The motivation for Model 2 is that the assumption of independent increments from Model 1 is a strong one which applies at *all* time scales. In particular infinite divisibility, a well known, property of Lévy processes [2], cannot hold in real systems due to physical constraints such as minimum packet sizes. On the other hand, independent increments can be a reasonable physical model in a range of scale above the level of packet size, but below the long-range dependence operating at, say, 1 second and above. Hence the fact that there is a typical time scale, namely t, where Model 2 holds, makes physical sense.

In Model 2 however, even if the law of $A[0,t)$ can be obtained, this is not sufficient in general to infer the complete law of A. In this case there *are* limits to system identifiability. However the joint distribution of (B,C), if determined, still gives valuable information on A and the queueing process it creates.

4 From Inversion to Estimators

The inversion expressions described in the previous section can be used as the basis of cross traffic estimators. In this section we define a number of such, and investigate their fundamental properties. In particular, we explain the underlying tradeoffs between errors of different kinds, and how they inter-relate, and discuss the problem of available data. To better focus on these key issues, we initially use simple cross traffic processes in simulations designed primarily to investigate and illustrate, and gradually introduce greater realism, for example issues of estimator bias and variance, as we proceed.

As we learn more about the behaviour of the estimators, important details emerge which result in them being modified, becoming more complex, but with better performance. The modifications fall into two groups. First, they arise naturally from the need to address 'practical' issues, such as the setting of parameters values. These must ultimately be based on data, which lead to an additional level of randomness, and the estimators thereby take on a stronger adaptive character. Second, opportunities arise to propose refinements to the estimators through various kinds of hybridisation. This leads to improved performance but the additional complexity has it own drawbacks, including lack of tractability, and potentially a lack of robustness.

The final estimators we propose do not appear until the end of Section 4.2. The detailed examination of their performance is then given in Section 4.3, where we use burstier and more realistic cross traffic models, and offer more systematic performance results.

4.1 The Underlying Estimators

We construct the initial estimators in two steps. First, the observable conditional densities $f_r(s) = P(S = s|R = r)$ and $g_r(u) = P(S-R = u|R \geq r)$ are estimated. To do so, we simply use the empirical frequencies, denoted by $\hat{f}_r(s)$ and $\hat{g}_r(u)$ respectively, based directly on the observed (r,s) pairs in the data. (If there are no samples for a given $r = r'$, which is often the case even though $f_r(s)$ is typically positive, we set $\hat{f}_{r'}(s) = 0$ for all s except the largest where we set the density to 1. If there are none for all $r \geq r'$, then $\hat{g}_{r'}(u) = 0$ for all u, except the largest.) Such estimators are intuitive, and enjoy the property that they are naturally normalised. By this we mean that their empirical CDFs, $\hat{F}_r(s) = \sum_{i=0}^{s} \hat{f}_r(i)$, and $\hat{G}_r(u) = \sum_{i=x-t}^{u} \hat{g}_r(i)$, monotonically increase from 0 up to 1, as a CDF should.

In the second step, we select inversion expressions from the previous section, and replace each of the exact observables $f_r(s)$, $g_r(u)$ and $F_r(s)$ by their estimated counterparts.

The first estimators of $c(l)$ we consider are defined in Equations (34) through (36). The symbol $\mathbf{r}$ is used for the free parameter rather than r, to avoid confusion with the latter's use as a sample of the excess delay variable R. Recall that since l is fixed when estimating $c(l)$, specifying $\mathbf{r}$ is equivalent to setting $\mathbf{k} = \mathbf{r}+l+x$, defining the corner $(k^*,l^*) = (\mathbf{k},l)$ of the leftmost angle used by the estimator.

Estimator $\hat{c}_1$ arises from Equation (19). It differs from that equation however in that $\mathbf{r}$ is not set to $t-x$, corresponding to the largest range of r under Class 1, but is free to take any value both below and above $t-x$. It can therefore be seen to be of either Class 1 or 2, depending on the situation. We have encountered its Class 2 form already in Equation (25).

Estimator $\hat{c}_2$ arises from Equation (24). Again, we do not prescribe the value of the parameter $\mathbf{r}$, but allow it to 'operate' as either Class 1 or Class 2. In the latter case, it can be seen as a form of Equation (18) where the weights are uniformly chosen, and can be naturally contrasted to estimator $\hat{c}_1$ using the same $\mathbf{r}$ value.

Estimator $\hat{c}_3$ arises from Equation (22). It is similar to $\hat{c}_2$ but with the addition of correction terms. Unfortunately, these terms have a negative impact on estimation performance, as we show below.

$$\hat{c}_1(l) = \hat{g}_{\mathbf{r}}(l+x). \tag{34}$$

$$\hat{c}_2(l) = \frac{1}{N}\sum_{r'=\mathbf{r}}^{\mathbf{r}+N-1} \hat{f}_{r'}(r'+l+x). \tag{35}$$

$$\hat{c}_3(l) = \frac{1}{N}\sum_{r'=\mathbf{r}}^{\mathbf{r}+N-1} \hat{f}_{r'}(r'+l+x) + \frac{\hat{F}_{\mathbf{r}}(\mathbf{r}+l+x-1) - \hat{F}_{\mathbf{r}+N}(\mathbf{r}+N+l+x-1)}{N}. \tag{36}$$

Estimators 2 and 3 use $r = \mathbf{r}$ to $r = \mathbf{r}+N$. To ensure that, for a given $\mathbf{r}$, they use the same amount of information as $\hat{c}_1$ in order to simplify comparison, we take N large enough so that each uses all observations of $R \geq \mathbf{r}$. This also eliminates the need to perform parameter selection with respect to N. It is important to note however that the selection of N has now effectively become random, which complicates any formal analysis. This is the first of many examples of where a practically sensible means of parameter selection in fact corresponds to the creation of more usable, but more complex estimators.

We now define our estimator for $h(l,k)$. We must distinguish between points in the exclusion zone and those that are not. Outside the zone, that is for $k-l-x \geq 1$, the estimator arises from

Equation (33) and is

$$\hat{h}(k,l) = \hat{F}_{k-l-x}(k) + \hat{F}_{k-l-x}(k-1) - \hat{F}_{k-l-x+1}(k) - \hat{F}_{k-l-x-1}(k-1). \quad (37)$$

The estimator for the horizontal aggregates in the exclusion zone follows from Equation (20):

$$\hat{c}(l+x,l) = \hat{F}_0(l+x) - \hat{F}_1(l+x). \quad (38)$$

Although it is beyond the scope of this paper to examine in detail estimators in Class 3, for completeness we show how they can be defined. Estimators of the Lévy exponent $\alpha(u)$ can be defined using Equation (30) directly:

$$\hat{\alpha}_0(u) = \frac{-1}{t} \log \hat{E}[e^{-uA(t)}]. \quad (39)$$

$\hat{E}[e^{-uA(t)}]$ can be set to $E[e^{-u\hat{A}(t)}]$ and computed using any of the above estimators for $A(t)$. Alternatively, we can use the principles of Class 1 inversion directly in the continuous time and space framework. The queue behaves linearly during $[T_n, T_{n+1}]$ if $R_n \geq t - x$, where $A(t)$ is simply $R_{n+1} - R_n + t - x$. Since $A(t)$ is conditionally independent of R_n, we can estimate $E[e^{-uA(t)}]$ as the sample mean of $e^{-u(R_{n+1}-R_n+t-x)}$ conditioned on $R_n \geq t - x$:

$$\hat{\alpha}(u) = -\frac{1}{t}[\log \sum_{n \in L_t} e^{-u(R_{n+1}-R_n+t-x)} - \log |L_t|], \quad (40)$$

where $L_t = \{n | T_{n+1} - T_n = t, R_n \geq t - x\}$. Since $\hat{\alpha}(0) = 0$, this estimator is also naturally normalised. We can therefore perform an inverse Laplace transform to recover, in fact define, an estimator $\hat{c}_4$ of the density $c(l)$:

$$\hat{c}_4(l) = \mathcal{L}^{-1}\hat{\alpha}(u). \quad (41)$$

Again, one could also use this Class 1 inspired estimator in a Class 2 manner, at the cost of increased bias.

4.2 Properties & Refinements

We use simulations with simple arrival processes, using a minimum number of parameters, to illustrate important factors affecting the estimators above. The properties discussed however are generic and remain valid in more general settings.

As with the analysis of Section 2, the system is fully discretised, with slotted time corresponding to the transmission time $\delta = 8d/\mu$ [sec], where $d = 10$ is the size of the slot measured in bytes, and μ is the output link capacity. In this setting Poisson arrival streams correspond in fact to i.i.d. time series with marginals being Poisson random variables with parameter $\lambda\delta$, and the packet service time, queue system time, and delay values are integer multiples of δ. However, a $d = 10$ byte granularity means we can think of the unperturbed system as M/G/1 for most practical purposes, whilst reducing computational issues and estimation variance. In this section and the next it will be convenient to present results either as integers from the discrete time system, l, k, r ecetera (already normalised by δ), or in units of bytes.

We use periodic probing streams with probes of size $p = 40$ bytes, so $x = p/d = 4$, with period $t = 10p/d = 40$ slot units, or 400 bytes. Cross traffic arrivals are taken to be 'Poisson', also with constant packet size p bytes. The above cross traffic and packet size combination corresponds to a particularly simple example of a measure A with stationary and independent increments. More realistic packet size distributions will be considered later in this section, in particular when evaluating performance in Section 4.3, and more complex arrival processes in Section 5.

4.2.1 The Issue of Available Data

To understand how the estimators behave, it is essential to know the environment they operate in. The following paragraphs examine this in detail for the system studied in this subsection, characterised by the parameters above and $\rho = 0.8$ and $\delta = 0.25$ [ms].

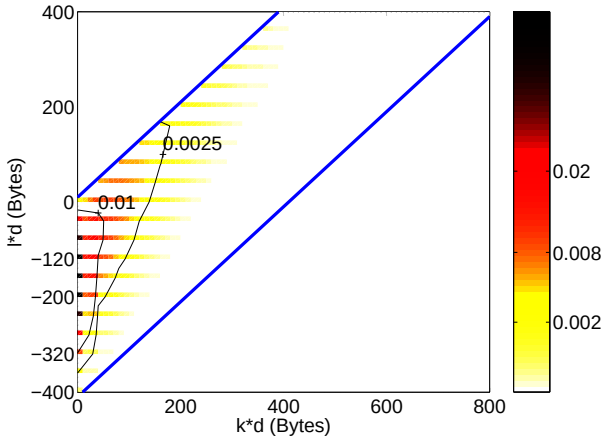

Figure 3: The density $h(k,l)$ of (B,C) for $\rho = 0.8$ (darker tones indicate higher density), and corresponding contour lines giving probability per 'pixel'. The density is concentrated on discrete l values corresponding to whole numbers of packets. The two values of l used in Figure 7, $-120/d$ and $-320/d$, are shown (d is the number of bytes per time slot).

We begin with Figure 3, where the shaded area visualises the joint density $h(k,l)$ of (B,C). The density is concentrated on lines corresponding to whole numbers of packets, and lies away from the lower edge of the strip for almost all l values. This indicates that the weak and strong assumption will hold for many corners (k^*, l^*) well inside the strip. The density is particularly concentrated near the l axis, corresponding to low per-slot burstiness of the cross traffic. The superimposed contour lines give an idea of the probabilities corresponding to the 'pixels' of this shading, which were drawn at full slot resolution[1]

Knowing where the density becomes negligible tell us which r values are necessary to measure it, and therefore informs the choice of the parameter value **r** which controls the range of r used by the estimators. The next question is, how available are these desirable **r** values? or equivalently, what is the probability

[1]For visual clarity, the contour lines, here and elsewhere, were smoothed to emphasise the l values corresponding to whole numbers of packets. At other l values the contours cut in much closer to the l axis.

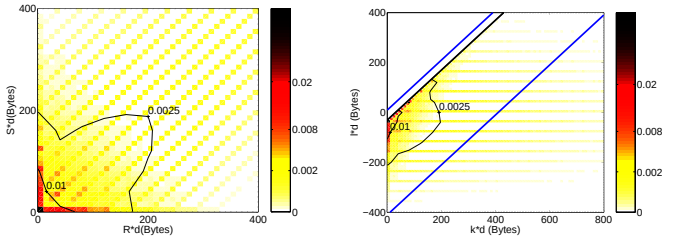

Figure 4: The density $m(r,s)$ of (R,S) (left), and its transformation into *angle density* $a(k,l)$ (right). Darker tones indicate higher density. Contour lines are for probability per 'pixel'. Lines of constant $u = r - s$ are mapped to horizontal lines in the (k,l) plane.

that the angle sets they correspond to will be seen? It is instruc-

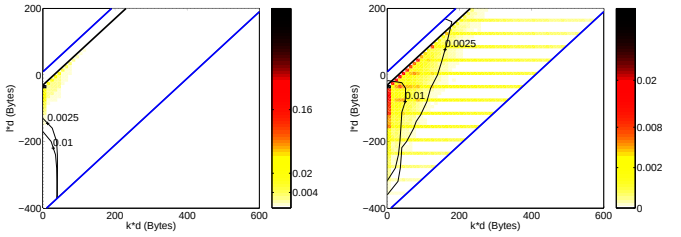

Figure 5: Superimposing angle density onto contours of $h(k,l)$ for $\rho = 0.2$ (left) and $\rho = 0.8$ (right). The angle density is given by the shading, and the cross traffic density by two contour lines. The degree of coverage of h by the angle density varies significantly.

tive to first visualise the density $m(r,s)$ of (R,S), as seen in the left plot in Figure 4. Mass is concentrated on lines of constant $u = r - s$ at large r, since there the queue cannot empty between the consecutive probes corresponding to r and s (recall that their separation is fixed here at t), and so they must share a busy period. The probe separation u is then constrained to be multiples of a cross traffic packet service time. Note that that marginals of R and S are identical.

To see how the density $m(r,s)$ impacts on estimation, it is in fact more useful to transform this information on 'available data' into a form which is directly readable in the (k,l) plane. Recall that there is a 1-1 mapping between (r,s) pairs and angle corners: $(k^*,l^*) = (s, s-r-x)$. Applying this mapping to $m(r,s)$ induces what we call the *angle density*, $a(k,l)$. Figure 4(b) displays the angle density, together with corresponding contour lines. It allows us to directly see where angles are likely to lie in a given experiment. The affine mapping has taken vertical/diagonal/horizontal lines in (r,s) space and mapped them to diagonal/horizontal/vertical lines in (k,l) space respectively.

Figure 5 gives a representation of angle density $a(k,l)$ and the cross traffic density $h(k,l)$ in the same plot, for two values of ρ. To avoid overcrowding the figure, the shading is given for angle density with no contours, and two contours are given for h, with no shading. As is clearly seen in the low utilisation case, where h is concentrated is not necessarily where the angle density is located. More generally, it is clear that to resolve h

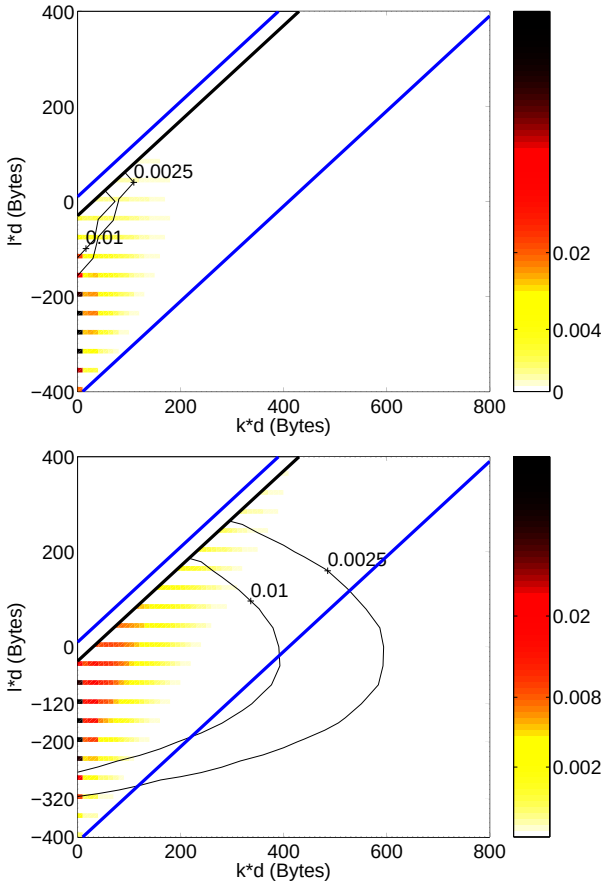

Figure 6: Superimposing contours of the 'available mass' used by estimators, over the density $h(k,l)$, for $\rho = 0.4$ (top) and $\rho = 0.8$ (bottom).

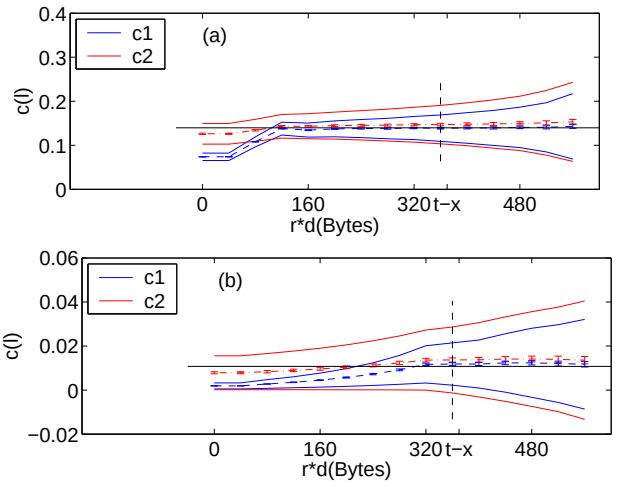

Figure 7: Estimator bias and variance as a function of **r**. The horizontal line is the true value of $c(l)$. (a) $l = -3p/d$, or -120 bytes. The bias begins at -80 bytes. (b) $l = -8p/d$, or -320 bytes. The bias begins at -280 bytes.

well across the (k,l) plane, we need a sufficient 'coverage' of angle density, and that whether this is achieved will depend on the number of observations as well as the queueing statistics

(which are a function of ρ, or more precisely, of the combined cross traffic and probe traffic processes).

Angle density shows where the 'available data' is located in the strip, however estimators make use of *sets* of angles. Thus, to see what is actually available for an estimator to use, we must sum over these sets. Recall from the previous section that each choice of l and $\mathbf{r}$ designates a set of angles whose corners can be defined by l and constant $\mathbf{u} = s - \mathbf{r} = l + x$, or equivalently $\mathbf{k} = \mathbf{r} + l + x$. The mass contained in these angles, that available to an estimator, is precisely the integral of $m(r, s)$ over the subset (see for example Figure 2(a)) defined by $s - r = \mathbf{u}$ for $r \geq \mathbf{r}$, which is also precisely $g_{\mathbf{r}}(\mathbf{u})P(R \geq \mathbf{r})$. In other words, this is simply the sum of the angle density $a(k, l)$ contained in the horizontal segment defined by a given fixed l and $k \geq \mathbf{k}$. The lower plot in Figure 6 repeats the $h(k, l)$ density from Figure 3, this time without the contours. Instead, contours are drawn based on the density of *available mass* used by an estimator as just defined. Using them, for any given l we can easily see which regions in the strip are data rich or data poor from the point of view of an estimator of $c(l)$. The answer clearly depends on l. The upper plot in the figure shows the two densities for $\rho = 0.4$, where the coverage is worse.

In conclusion, two important questions: whether strong or weak assumptions hold (and the shape of these regions), and whether there is sufficient data available to an estimator, depend strongly on l. Furthermore, the interplay between the densities $h(k, l)$ and $a(k, l)$ over the plane is crucial. The feasibility of estimation depends strongly on l, and furthermore there is an intrinsic difficulty in that the degree of 'coverage' may not be adequate. For low levels of burstiness/utilisation, the marginal of R is concentrated near $R = 0$, resulting in available mass which is strongly concentrated near $(k, l) = (0, 0)$. However under these same circumstances, $h(k, l)$ is concentrated near $(k, l) = (0, -t)$. The overlap of the two is small, and any estimator will have great difficulty, essentially because the region where the data is needed in order to measure the (b, c) values which occur, is precisely where data is scarce. Coverage is determined not by utilisation alone but by the spread of (r, s) values seen, which depends on ρ, burstiness, as well as the number of observations.

4.2.2 A Bias/Variance Tradeoff at Fixed l

We now examine estimator performance for $l = -3p/d$, or -120 bytes. For this value of l, the lower plot in Figure 6 show that conditions are good: the available mass contour lines shows that the region where $h(k, l)$ is concentrated is well covered, and also that there is considerable mass available in the angles both inside and to the right of the strip. Furthermore, $h(k, l)$ is small for a considerable distance to the left of the righthand boundary of the strip. Hence both $\hat{c}_1$ and $\hat{c}_2$ can be expected to perform well either as Class 1 or Class 2. Figure 7(a) compares the mean and standard deviation of the estimators, based on $n = 1000$ probes, as a function of the parameter $\mathbf{r}$. The mean and standard deviation were estimated using $N = 1000$ independent experiments, each yielding a single sample for each estimator (and for each $\mathbf{r}$). The estimated mean is shown with 1.96σ confidence intervals near the center of the plots. The outer curves are drawn one standard deviation (of the estimator) to either side of the means.

For $r \geq t - x$ each estimator operates as Class 1. As expected each gives approximately unbiased estimates of $c(l)$ in this case. Also has expected, the uniform weighting scheme of $\hat{c}_2$ is less effective, resulting in larger variance. The correction terms in Equation (36) separating $\hat{c}_2$ and $\hat{c}_3$ identically cancel under Class 1 in theory, however estimates of them do not. Effectively an imperfect estimate of zero is added! resulting in increased bias and variance (not shown).

For $r < t - x$ each estimator operates as Class 2. As $\mathbf{r}$ decreases, we expect each to become biased. This is indeed what is observed, however there is **no sharp change** at $r = t - x$, since the strong assumption holds very well. For example a bound on the total error due to the assumption: mass ignored plus the undesirable mass included in the angle at $(\mathbf{k}, l)$, is only 0.1% of $c(l)$ (i.e. $(c(l) - c(\mathbf{k}, l) + b(\mathbf{k}, l - 1))/c(l) = 0.001$). At small $\mathbf{r}$ however when the strong assumption finally fails, the bias of $\hat{c}_1$ is much worse, because its weighting scheme was not designed to cope with the errors inherent in Class 2. As before however, the variance of $\hat{c}_1$ is lower, as greater weight lies in the data rich zones where $\mathbf{r}$ and s are smaller. Again the correction terms of $\hat{c}_3$ worsen its performance (not shown) relative to $\hat{c}_2$. Since they are in the form of a difference of two quantities of similar size, they are sensitive to errors.

In conclusion, there is a classic bias variance trade off operating, which begins once the strong assumption ceases to hold. This value of $\mathbf{r}$, which plays the role of the *effective* Class 1/Class 2 boundary, is clearly visible in Figure 7(a) as the point where the bias of $\hat{c}_1$ begins to be noticeable. To the left of this point, c_1 has rapidly decreasing variance at the cost of increased bias, whereas $\hat{c}_2$ reacts more slowly. Figure 7(b) shows an analogous study for a smaller $l = -8p/d$, or -320 bytes. Similar conclusions on bias and variance hold as before for each estimator separately, however the point marking the beginning of the bias increase now occurs earlier at $r = -8p/d$, again in line with the failure of the strong assumption as seen visually in Figure 6 (lower plot).

Consider now the effect of reducing n on the results in Figure 7. To first order, the qualitative behaviour remains the same, but the standard deviation of estimates increase. To mitigate this, one could be led to select smaller values of $\mathbf{r}$ to capture more data, and in order to control the resulting increase in bias, c_2 may becomes more attractive in comparison to c_1. Thus which estimator is preferable (at least in terms of bias), and in particular the success of a Class 1 inspired approach, is also dependent on the global amount of data available, and not only its repartition with l.

The above demonstrates the value of the Class 2 inspired estimator, and the necessity on not insisting on *always* having linear behaviour of the queue for estimation. This is a key advantage of the techniques we present here, which, being based on (R, S) pairs arising from adajent probes, can be though of as falling into the packet-pair class in a generalised sense. In earlier packet pair methods it was crucial that the probes share the same busy period **and** were back to back. Here this is not the case. Class 2 based estimators not only can be used, but they may even outperform those based on Class 1 ideas. Furthermore, because of

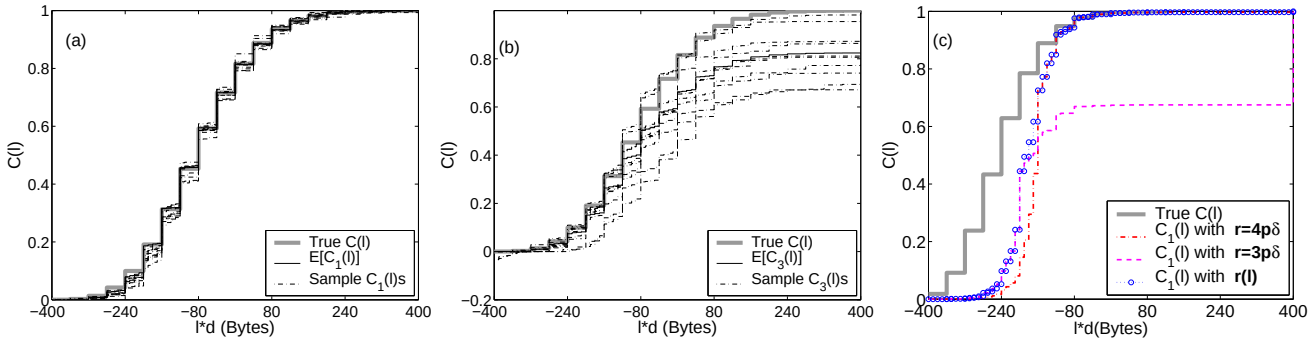

Figure 8: Comparison of CDF estimates (solid dark line) and true CDF $C(l) = P(C \leq l)$ (thick grey line). (a) $\hat{C}_1$ with $\mathbf{r}(l) = 4p/d$, expectation and 8 samples, bias and variance are small with $\rho = 0.8$ but depend on l; (b) $\hat{C}_3$, with $\mathbf{r}(l) = 4p/d$, estimates are individually worse and not naturally normalised; (c) $\hat{C}_1$, three different choices of $\mathbf{r}(l)$. Varying $\mathbf{r}$ with l performs better than either of the best static choices $\mathbf{r}(l) = 3p/d$ and $4p/d$.

this, greater values of t can be used than before, because the push into the data-poor regime, to the detriment of Class 1 estimators, is no longer a fundamental barrier. Finally, the need for Class 2 is in any case generic, since the data-poor regime cannot be avoided when estimating the distribution $C(l)$, since the estimation will always be data-poor *for at least some l*.

The innovation of our approach can be described as follows. In the past delay observations were divided into two categories according to an (inferred) busy period/idle period criterion for probes [17]. Although this is a very intuitive approach well suited to heuristic methods, it is difficult to carry it further. Instead, we employ conditioning with respect to R, thereby providing a sequence of subsets of probe delay, indexed by r. These subsets vary in their proportions of probes pairs which share, or not, the same busy period. From each subset an estimate can be obtained. For large enough r, the sets fall into Class 1 and therefore contain only the busy cases, so that estimates based on these are bias free. As we enter Class 2 at smaller r, this is no longer the case but it remains approximately so until the strong assumption is broken. By viewing the problem in this way, we have a sequence of estimators of steadily decreasing bias but increasing variance as r increases. The estimation problem can now be framed as an optimisation problem, whereby the point along the sequence from which we are prepared to accept estimates is decided based on some metric combining bias, variance, and some measure of cost. This may involve a notion of 'probe budget', or a fixed time horizon for measurement together with some invasiveness constraint.

4.2.3 From Density to Distribution

In this section we consider in more detail the dependence of the estimators on l. It is instructive to do this by looking at estimates of the CDF $C(l)$, defined by

$$\hat{C}_j(l) = \sum_{i=-t}^{l} \hat{c}_j(i), \tag{42}$$

for each of $j = 1, 2, 3$, rather than examining the density estimates $\hat{c}_j(l)$ directly. To complete the above definition, we must also specify the parameter values $\mathbf{r}(l)$ used for each l. The estimators can be thought of as a random functions, whose samples are the empirical CDFs. The expection, bias, and variance of each estimator are likewise functions of l. Note, by the definition above and the definitions of the CDF of the observables given earlier, that $\hat{C}_1(l) = \hat{G}_{\mathbf{r}}(l + x)$.

Figure 8(a) compares $\hat{C}_1(l)$ with $\mathbf{r}(l) = 4p/d$, again based on $n = 1000$ probes, against $C(l)$, using the same cross traffic as in Section 4.2.2, and again using $td = 10p$ bytes. The expectation function (calculated as the average of $N = 1000$ realisations) is very close to the true CDF, and the variance, illustrated informally by the plotting of 10 individual samples, is likewise small. Both however are functions of l: the bias is greatest at small l, whilst the variance is larger at intermediate values. The lower bias at larger l is easy to understand from Figure 3 given how the diagonal $\mathbf{r}(l) = 4p/d$ moves to the right of where $h(k, l)$ is significant, allowing the strong assumption to hold. Note that, since $\hat{c}_1(l) = \hat{g}_{\mathbf{r}}(l+x)$, and $\mathbf{r}(l)$ is constant, the natural normalisation of $\hat{G}_{\mathbf{r}}(u)$ is naturally transferred to $\hat{C}_1(l)$. Thus, even though estimates of $\hat{c}_1(l)$ must eventually be poor when l is very large, this does not prevent good behaviour of the CDF. Thus natural normalisation is an extremely desirable property.

Figure 8(b) offers exactly the same comparison as in plot (a), only for $\hat{C}_3(l)$. Because of the correction terms in Equation (36), $\hat{c}_3(l)$ does not possess the natural normalisation enjoyed by both $\hat{c}_1$ and $\hat{c}_2$. Consequently, the errors in the density (apart from being individually considerably worse as discussed above) are not constrained to cancel at large l in the same way, leading to CDF estimates with fundamentally flawed properties. As a result, $\hat{c}_3(l)$ will not be considered further.

Since $\rho = 0.8$ in Figure 8(a), there is enough data to provide good estimates at most l values of significance. Figure 8(c) shows how the performance of $\hat{C}_1(l)$ drops significantly when $\rho = 0.4$ (expectation estimates only are shown, again based on $N = 1000$). The l dependence of the bias is now strong and clearly visible, as is the dependence on the choice of $\mathbf{r}$. When moving from $\mathbf{r} = 3p/d$ to the lower diagonal of $4p/d$, the bias becomes even worse at small l but improves markedly at large l. This suggests that an 'adaptive' strategy, where $\mathbf{r}(l)$ truly depends on l, could be used improve estimation. An example of this is given, where a transition from $\hat{C}_1(l)$ with $\mathbf{r} = 3p/d$ to

$\hat{C}_1(l)$ to $\mathbf{r} = 4p/d$ occurs at $l = -4p/d$ (i.e. at $A(t)d = 6p$ bytes or 6 packets arriving between probes).

As the potential benefit of making $\mathbf{r}$ a function of l is significant, we next examine this issue in detail. We find that we are not free to adapt $\mathbf{r}(l)$ in any arbitrary manner, but must proceed carefully, and furthermore that the attempt to optimise performance in this way brings with it some intrinsic difficulties which negate some of the advantages.

From this point on, we omit results for $\hat{C}_2$, concentrating solely on $\hat{C}_1$. We do this mainly because the results for the two are very similar in the cases of interest, namely where available data is poor and where good estimator performance is a challenge. This is because in that case typically either 0 or 1 angles are found at any given point in the strip. Consequently, the weights appearing in the definition of $\hat{c}_1$ become, in a sample path sense, uniform, just like those of $\hat{c}_2$. Another way of stating this is that the empirical estimate of $g_r(u)$ is so poor that it completely fails to reproduce the features of the true distribution. The other reason is that, in cases where more data is available and the empirical estimates are not so poor, the best performing estimates (based on the results we show later and others not shown in this report) are those which emphasize the avoidance of bias. In the language of Figure 7, the region at small l where the c_1 and c_2 become significantly different is avoided.

4.2.4 Determining $r(l)$

In this section we examine issues relating to a definition of $\mathbf{r}(l)$, and propose an algorithm to estimate it in practice. The extra steps required to build a composite estimator, in the spirit of Figure 8(c), which uses $\mathbf{r}(l)$ to direct which of a bank of available constant $\mathbf{r}$ estimator to use at any given l, are left to section 4.2.5.

Our guiding principle follows from the observations of the previous section: if $\mathbf{r}(l)$ can be chosen to match where the strong assumption begins to fail, essentially tracking the boundary where the density in Figure 3 drops off, then the bias-variance tradeoff observed in Figure 7 could be well managed for each l. However, this approach will fail when there is insufficient data to measure the position of this boundary, as for example in Figure 5(b), where the density $h(k,l)$ of (B,C), roughly speaking centered about $(k,l) = (0,-t)$ (note the tiny black region), is well separated from the density of available data, centered about $(k,l) = (0,-x)$, corresponding to $(r,s) = (0,0)$, i.e. with high probability the probe delays are close to the minimum. In fact there are a number of interconnected issues here which must be considered before a suitable choice of $\mathbf{r}(l)$ can be found, as we now detail.

Redefining the strong assumption

The density $h(k,l)$ is two-dimensional, and is therefore intrinsically difficult to estimate well. In particular, attempting to estimate quantiles of $c(l)$ for each l individually may be a hopeless task when there are only a small number of angles, perhaps even zero, available at that l. Consequently, we will redefine the strong assumption in terms of the CDF of $c(l)$, $C(l)$, to help stabilise estimation.

It is easy to see from Equation 8 that $C(l) = H(k,l)$ for $k \geq l + t$. We approximate this by $H(\mathbf{k},l)$, $\mathbf{k} \leq l + t$, thereby potentially ignoring the part of the density nearest the right hand edge of the strip. This creates an error

$$e_s(\mathbf{k};l) = H(l+t,l) - H(\mathbf{k},l) \geq 0 \tag{43}$$

corresponding to the sum of $h(k,l')$ over a trianglar region defined by $k > \mathbf{k}$, $l' \leq l$, and the lower boundary of the strip. Such a definition is similar in spirit, but technically slightly different, to what one would obtain if we assumed the single-l strong definition of Equation 23 over a range $l' \leq l$. A minor difference is that the new definition includes the density lying strictly below the corner $(\mathbf{k},l)$, which was previously excluded. The larger difference is that the previous definition implied that, for any l, the estimate of $C(l)$ would assume zero mass in a 'strong assumption region' adjacent to the lower edge of the strip for **all** $l' \leq l$. In contrast, here for any given l, the mass neglected is only that in the triangular region described, and does not enter $k < \mathbf{k}$. Thus the new definition demands progressively weaker conditions as l increases, compared to the previous one. In practical terms, it makes better use of the available data.

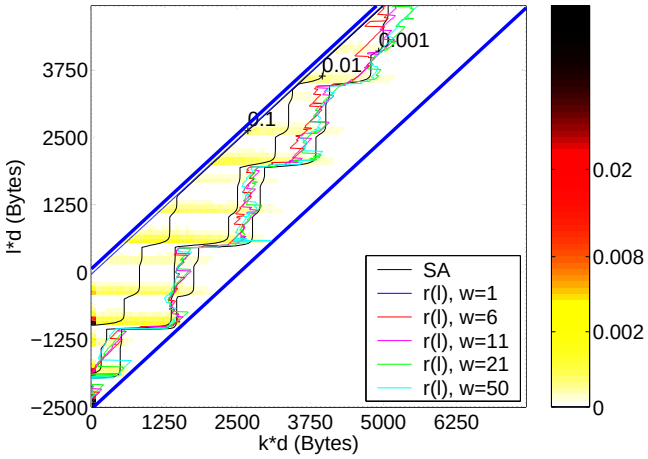

Figure 9: Examples of $C(l)$-based strong assumption curves for $\theta = \{0.0004, 0.01, 0.1\}$ (3 grey curves) for $\rho = 0.8$. Estimates from the saturation algorithm are also shown for different window sizes w.

We define the *strong assumption curve* at probability threshold θ as follows:

$$k_s(l;\theta) = \max(0, \arg\min_{k \geq 0} e_s(k;l) < \theta) \tag{44}$$

that is, for each l, the leftmost k value within the strip such that the error due to the strong assumption does not exceed θ. Figure 9 gives examples (estimates makes using $N = 1000000$) of the strong assumption curve for three threshold values at high utilisation. In this figure, and others below, we show results for a more realistic tri-modal distribution of packet size, introduced formally in Section 4.3.

Estimating the strong assumption curve

Recall that the sum over a rectangle set $H(k,l) = F_{k-l-x}(k)$, our approximation to $C(l)$, can be estimated using $\hat{F}_r(s)$, or $\hat{G}_r(l+x)$ as defined above (more precisely, from Equation 28, if $r = \mathbf{r}$ corresponds to θ under the new definition it follows that $G_{\mathbf{r}}(l+x) \leq 1-\theta$, since the rectangles in the sum (beyond

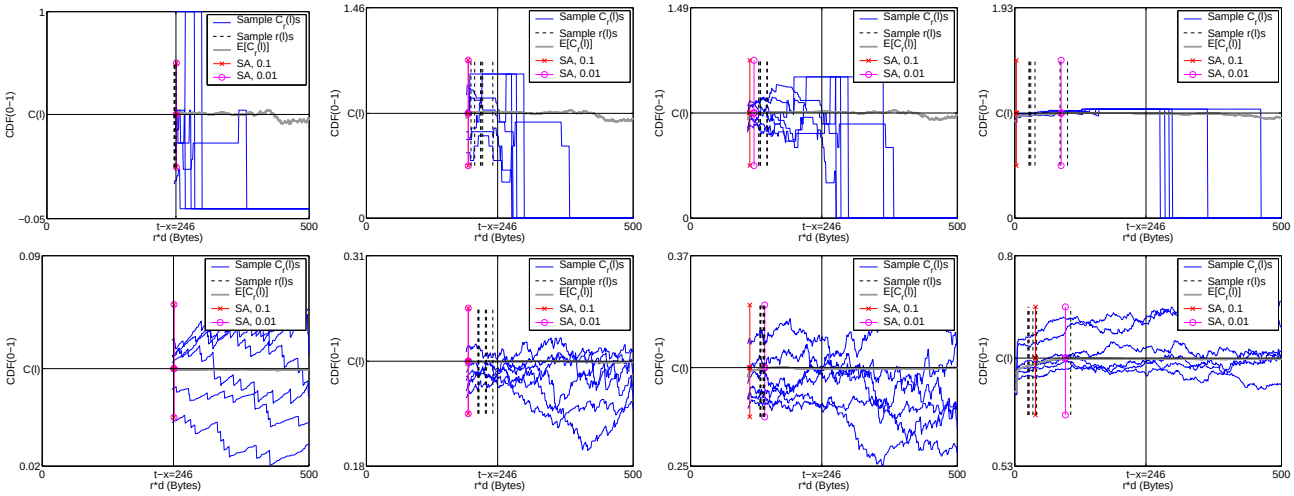

Figure 10: Strong curve estimation with $\rho = 0.2$ (top) and $\rho = 0.8$ (bottom) for (left to right) $l = \{4, 60, 140, 300\}$. The 6 thin lines (resp. single thick grey line) are estimates of $H(\mathbf{k}, l)$ using 500 (resp. 1 million) probes. The $\mathbf{r}$ values selected by the saturation algorithm are shown as vertical lines, can be compared to the strong assumption values $k_s(l;\theta)$ for $\theta = \{0.00004, 0, 01\}$.

the first term) have corners further to the right). Figure 10 plots $\hat{G}_r(l+x)$ against r (rd [bytes]) for a range of four l values, for low (top row) and high (bottom) utilisation. The noisy curves are several independent estimated functions each based on a single realisation of $n = 500$ probes, whereas for comparison, the thick grey curve derives from a realisation employing 1 million probes. As r increases, $G_r(l+x)$ increases monotonically to attain $C(l)$ (shown as the horizontal line) at $r = t - x = 246$ (the full height vertical line). The estimates roughly follow this pattern, however since the available data monotonically decreases with r, the crucial limiting behaviour becomes obscured by noise which can take extreme values. Moreover, the curves show non-ergodic features in that they oscillate about a limiting level which is not necessarily $C(l)$ but some random offset from it. As a result, it is not feasible to target the point lying on a strong assumption curve given by a small threshold θ. We therefore adopt a less ambitious approach which aims to find the point at which the steadily increasing phase of the estimate curve saturates. We first smooth the curve to reduce the sample variability, so that the systematic increase at small r can be seen more clearly. The intuition is that when the underlying 'expected' curve has saturated, then the variability will cause the curve to cease to become monotonic despite the smoothing.

Saturation Algorithm:

i) select a window size w (performance insensitive to value)
ii) smooth the $\hat{G}_r(l+x)$ estimates using a moving average window filter of width w (the filter is causal, thus there is an edge effect over the first $w-1$ values).
iii) if $r = t - x$, set $\mathbf{r} = t - x$ and exit, else set $\mathbf{r}$ to the first r for which the smoothed curve ceases to be non-decreasing[2].

The algorithm is guaranteed to terminate with a value $0 \le \mathbf{r} \le t - x$. Note that for $l < 0$ the minimum $\mathbf{r}$ value is constrained by the shape of the strip. In Figure 10 values are only plotted from the first entry into the strip on. The full height vertical line marks $r = t - x$.

As noted above, it is not feasible to locate the point $k_x(l;\theta)$ on a strong assumption curve for a given θ. By comparing the algorithm outputs in Figure 10, given by the thin half height vertical lines, against the thicker vertical lines corresponding to $k_s(l,\theta)$ for two values of θ, we see that the saturation algorithm outputs indeed do not track any particular θ value. Indeed, the algorithm is influenced not only by the distance to the saturation level $C(l)$, but also by the variability of the curves, which trigger the algorithm to exit once they become too severe in the downward direction. Thus in some informal sense, the algorithm is performing a tradeoff between bias and variance rather than being concerned solely with the strong assumption curve (which would correspond to emphasizing the bias only). Although this means that, unfortunately, the algorithm performance cannot be tested in isolation by comparing against a target θ, it is in other respects entirely appropriate, since as already noted, $k_s(l;\theta)$ may be inherently impossible to measure without bias if the coverage of h is poor. On the other hand, the expected $\mathbf{r}(l)$ curves (projected into the (k,l) plane) of Figure 9, estimated by using a million sample simulation, shows that the algorithm does on average output a function which roughly correspond to a strong assumption curve (in this case with $\theta \approx 0.001$) as originally intended. This graphs also illustrates the fact that $\mathbf{r}(l)$ generally decreases with l, following the strong assumption curve.

Figure 9 also shows the important property of insensitivity of the algorithm with respect to the window size parameter larger than 1. Here, d is 10 bytes and t is 250 slots. Hence, $w = 25$ (slots) corresponds to $t/w = 10$. Since the algorithm performance was good as long as w was neither too close to 0 or t, we use a default value of $t/w = 10$. Later, we also justify this choice by comparing performance of CDF estimation with various window sizes.

[2]In the implementation, it was not necessary to actually smooth, but only to see if the new window element entering on the right is smaller than the one departing.

4.2.5 Defining the Composite Estimator

They are many possible ways in which one could make use of $\mathbf{r}(l)$ to design new estimators. The example in Figure 8(c) at the end of Section 4.2.3 simply moved from one constant-**r** CDF estimator to another at a particular value of l. This simple approach in fact enjoys an important property which is not easy to guarantee in general. This is the fact that, since constant-**r** estimators are naturally normalised (i.e. they tend to 1 at large l), so is the new adaptive estimator, and this extends immediately to arbitrary $\mathbf{r}(l)$.

One can therefore define a $\mathbf{r}(l)$ based estimator as follows. First, calculate $\mathbf{r}(l)$ as in the previous section. For each of the different r values appearing in the function, calculate the corresponding estimator $\hat{C}_1(l)$. The idea is that by moving between members of this 'bank' of constant-**r** estimators, we can obey $\mathbf{r}(l)$ whilst simultaneously preserving natural normalisation. This property would **not** have been the case if instead for example we had tried to adapt the density estimate instead of the CDF, say by setting $\hat{c}(l) = \hat{g}_{\mathbf{r}(l)}(l+x)$.

The disavantage of the above naive or *raw composite* method is that there is no guarantee that the resulting CDF is monotonic, the CDF could move downward when we switch to a new member of the bank. Thus, although it could be used to estimate a given fixed quantile, it is not useful for measuring the probability of smaller sets, as it may assign negative probability to them. We investigated three methods which modify the above to form a *monotonic composite* estimator whose sample functions are both normalised and monotonic.

Monotonicity Algorithm:
First calculate a raw composite CDF $C(l)$. Then:
L2R: move left to right, forming $C'(l) = \max(C(l), C'(l-1))$.
R2L: move right to left, forming $C'(l) = \min(C(l), C'(l+1))$.
Data Pinning: obtain the number n_l of probes with $R \geq \mathbf{r}(l)$, initialize a set Q of processed l values to null. In order of decreasing size of n_l, recursively assign $C'(l) = C(l)$, then ensure its consistency (enforce monotonicity) at all l values already in Q, then add l to Q. More precisely: $\forall l' \in Q < l$, set $C'(l) = \max(C'(l), C'(l'))$ and $\forall l' \in Q > l$, $C'(l) = \min(C'(l), C'(l'))$.

The third algorithm 'pins' the estimate at the l values which have the most available data whilst enforcing monotonicity, proceeding recursively between the pinned values until the entire function is determined. This corresponds to a kind of constrained interpolation, weighted by available data.

Figure 11 gives an example of a raw composite estimate based on $\mathbf{r}(l)$, together with the three monotonicity enforcing algorithms just described. The upper curves are the expected CDF functions, obtained by averaging over $n = 1000$ independent experiments. The lower curves show the standard deviation of the same estimates as a function of l (using the same vertical scale). The raw curve shows a reasonably small bias at all l, and a clear lack of monotonicity (although this may be due estimation error: the expected curves may be monotonic though individual sample functions of course are not). In comparison, as expected L2R makes the estimate move up, and R2L move down. Perhaps unexpectedly, Data Pinning shows results which are almost indistinguishable from R2L. Since r(l) mostly

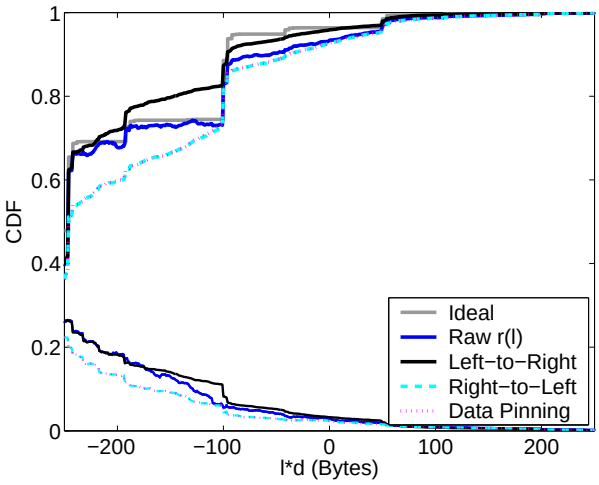

Figure 11: Statistics of composite estimators $\hat{C}'_1(l)$ based on $\mathbf{r}(l)$ with $t/w = 10$ ($t = 10$ and $\rho = 0.2$). The raw and three monotonic composite estimators are shown. Monotonicity causes a large bias (compare upper expected curves with the true CDF in grey) which is not compensated by a corresponding decrease in standard deviation (lower curves).

decreases with increasing l, R-t-L and data pinning algorithms perform close to each other. This is also true for the standard deviation of the two, which is better than that of raw and L2R. In all cases however, the monotonicity algorithms create significant differences in bias, for small changes in standard deviation. Bias is the main problem introduced by the need for an algorithm for monotonicity.

Three example sample paths of each estimator, in a data poor case with $\rho = 0.1$, are shown in Figure 12. Because their behaviour is similar, to reduce clutter we show Data Pinning but not R2L. The sample paths of the raw composite estimator are extreme. Typically they rise to 1 at small l where there is no data and hence where bias is extreme, before improving at intermediate l. At large l data is again scarce but the natural normalisation property limits the absolute size bias can take. The L2R estimator performs very poorly as it locks in the terrible performance at small l, estimates cannot decrease at larger l. Data Pinning (and R2L, not shown) perform much better, but we see that their variance is considerable. It is important to note that here we are zooming in performance under very difficult conditions where there is 'almost no data' for the estimator to work with. Under richer data scenarios, all these variants perform quite well.

From these results we learn that there is limited benefit from attemping to 'smooth' $\mathbf{r}(l)$, as a way of reducing the number of values that $r(l)$ takes and therefore the ill-effects of the monotonicity algorithm. Even if $\mathbf{r}(l)$ were taken to be piecewise contant with only two values, in data poor cases we may still be moving between sample CDFs which are very crude, resulting in large errors over large ranges of l values. Indeed, in scenarios where angles are so scarce that there are only j values of l where they can be found, the corresponding sample CDF will contain only $j-1$ jumps, a very crude approximation of the true $C(l)$ which has an uncountable infinity of them. This is

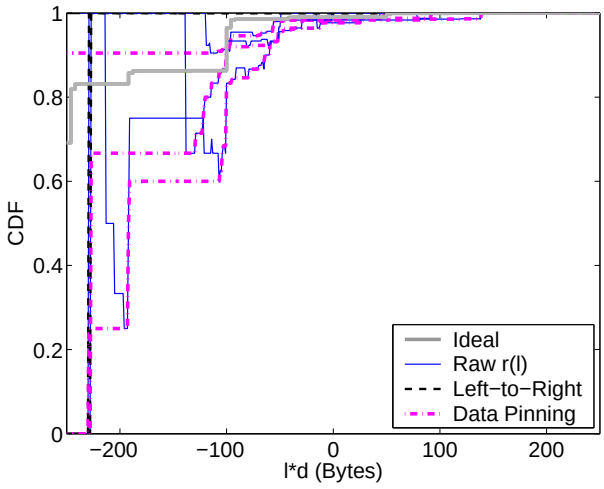

Figure 12: Sample Plots of Raw and Monotonized Estimates for $\rho = 0.1$ and $t = 1.67p_{max}/d$. We do not show the R-to-L estimates to reduce clutter. This plot illustrates how the bias and variance of per-l estimates are dependent. This dependence causes that across-l metrics (sup-norm and L1-norm) to have different trends than the per-l metrics.

typical of problems found in empirical estimation of discrete (or continuous) densities from limited data.

4.2.6 An Adaptive Constant r Estimator

The 'underlying' estimators have **r** a constant. They are attractive due to their simplicity, however in practice, one must select the value of the parameter **r**, and this must at some level be chosen from data to avoid very poor performance. In some cases one may already have a good idea of the important parameters controlling h, particularly ρ and t, and thereby have reasonable values of the marginal of R, from which appropriate values of **r** could be tabulated, for example by estimating some quantile of r_q of R. In other cases, quantiles of R could be estimated continuously over some timescale. At one extreme, **r** would be determined based only on the data used to estimate $C(l)$ itself, at which point the estimator can no longer be regarded as a constant **r** one, but a more sophisticated adaptive one where **r** is a random variable.

In the next section we consider an adaptive estimator of this type whereby **r** is selected according to the following principle: to locate the edge of available data. That is, it aims to find an **r** small enough so that some data will lie below it in the strip (as it is essential that the estimators will have some data to work with), but not to go much smaller (higher in the strip) than that, in order to avoid bias. The appropriate way to measure 'data available' is absolute in this case, rather than relative. Accordingly we choose $\mathbf{r} = \min\{t - x, r_*\}$, where $r_* = \hat{F}_r^{-1}((n-m)/n)$ is the r corresponding to having at least m observations. Note that when data is plentiful **r** will default to $\mathbf{r} = t - x$, as as r values to the right of the strip are always bias free.

There is still a need to automatically select m. We do not consider this here but examine performance using a range of different m values to determine the potential of this class of estimator.

4.3 Estimator Performance

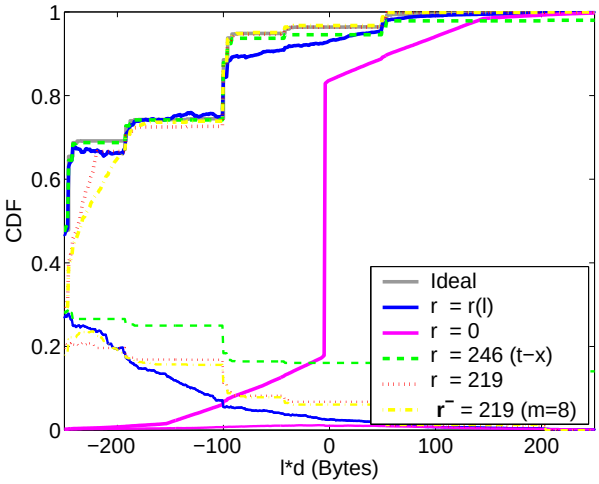

Figure 13: Examples of expectation and standard deviation of estimate functions for $\rho = 0.2$. Here $\mathbf{r}(l)$ is raw monotonic.

In this section we examine the performance of variants of $\hat{C}_1$, as defined at the end of the previous section, as a function of cross traffic, and of the probe periodicity t. Specifically, the variants and parameter values are:

- Constant **r**:
  - $\mathbf{r} = 0$, the naive packet pair heuristic
  - $\mathbf{r} = t - x$, pure Class 1 (low bias, high variance)
  - **r** = quantile corresponding to $m = 8$
- Adaptive constant **r**:
  - $m = 2$ use very little data (low bias, high variance)
  - $m = 8$ compare adaptive to constant above
  - $m = 50$ use more data (higher bias, lower variance)
- Composite estimator using $\mathbf{r}(l)$:
  - Data Pinning: the main candidate ($t/w = 10$)
  - raw: best case for composite method ($t/w = 10$)
  - R2L and L2R: for robustness comparisons

Note that, in particular under Model 1, we are not restricted to periodic probes, but it is convenient to continue to consider this case here so as to maximize the number of samples for a given number of probes n, and concentrate on the estimator performance as such.

We abandon constant packet sizes entirely from now on, and move to a trimodal packet size distribution similar to that found in the Internet:

$$p(i) = \begin{cases} 0.5, & i = 40; \\ 0.1, & i = 580; \\ 0.4, & i = 1500; \\ 0, & \text{otherwise}, \end{cases} \tag{45}$$

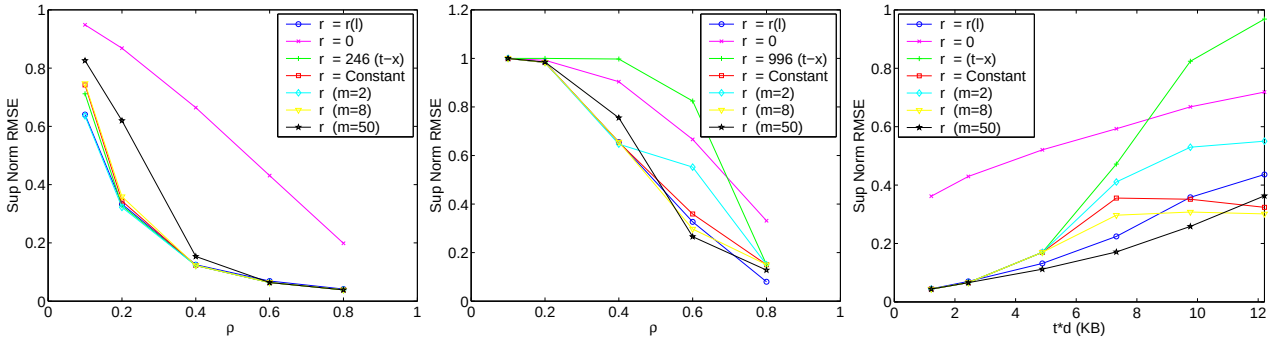

Figure 14: Sup-norm performance using trimodal packet sizes. (a) 7 estimators as a function of ρ, $t = 1.67 p_{max}/d$; (b) The same estimators with $t = 6.67 p_{max}/d$; (c) Dependence on t, with $\rho = 0.6$. Here $r(l)$ denotes the *raw* composite estimator.

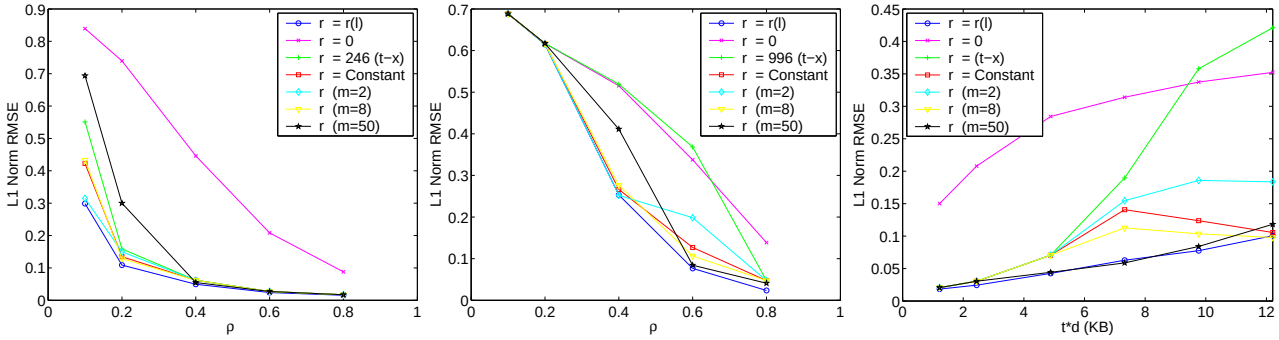

Figure 15: L1-norm performance using trimodal packet sizes. (a) 7 estimators as a function of ρ, $t = 1.67 p_{max}/d$; (b) The same estimators with $t = 6.67 p_{max}/d$; (c) Dependence on t, with $\rho = 0.6$. Here $r(l)$ denotes the *raw* composite estimator.

and use probes of 40 bytes (conveniently, each value of i is an integer multiple of $d = 10$). Although packet arrivals remain Poisson, the workload arrival process, which in an interval t consists of a Poisson distributed number of packets, each with random size given by the discrete density $\{p(i)\}$, is a Compound Poisson distribution, which is much more bursty when $p(i)$ is not concentrated on a single value (the constant packet size case). We calculate the CDF of this distribution by using a numerical approximation of an exact formula for $c(l)$, being simply the Poisson weighted (with parameter λt) sum of terms, where term j is the j-fold convolution of the density $p(\cdot)$ above, calculated directly in the time domain. The pointwise error can be controlled and was chosen here to be 1e-5, negligible compared to other factors.

We begin with Figure 13, where a similar representation to Figure 11 is given. We see that the naive $\mathbf{r} = 0$ estimator has bias so high that its variance function is small, indicating that the great majority of estimates share the same poor behaviour. Using $\mathbf{r} = t - x$ in this case produces very low bias but high variance, as expected. By entering into the strip and using $\mathbf{r} = 219$ (the quantile corresponding to $m = 8$ on average), we add bias at small l, but gain reduced variance over all l as a result. The adaptive version of this estimator, using $m = 8$ in a per-estimate sense, improves the bias performance with no variance penalty. Finally, the raw composite estimator shows low bias for most l values, and lower variance at most l values, indicating that it is worth pursuing estimators of this type.

Performance results of the type shown in Figure 13 are too detailed to allow coverage of the parameter space governing cross traffic characteristics. To assess estimator performance in a way which combines bias and variance, and examines an entire CDF, we define the following two measures, each of which returns a single number to evaluate a given sample function.

- Sup: $\mathcal{E} = \sup_l |\hat{C}(l) - C(l)|$
- L1: $\mathcal{E} = \frac{1}{l_q + t + 1} \sum_{l=-t}^{l_q} |\hat{C}(l) - C(l)|$,

where l_q the qth quantile of $C(l)$.
The first of these measures the worst departure from the true CDF over all l, whereas the second gives a measure of the average departure. We cannot let $l_q = \infty$, as this would be identically zero for any two distributions, no matter how different, due to the domination of the tail where $C(l) \approx 1$ out to infinity. Instead, we assess the degree of difference only over the main body of the distribution. In practice we use $q = 0.95$.

For each measure the random variable $\mathcal{E}$ takes values in $[0, 1]$. Our performance metrics are the MSE, defined as

$$MSE = \mathbb{E}[\mathcal{E}]^2 + \mathrm{Var}[\mathcal{E}], \tag{46}$$

of the corresponding measures. These also take values in $[0, 1]$. The expection and variance of $\mathcal{E}$ are estimated, in the usual way, using $N = 1000$ independent experiments, and are of interest in their own right, as components of the MSE.

Results are given in Figure 14 for the (root) MSE using the Sup measure. The three plots sample the (ρ, t) parameter space

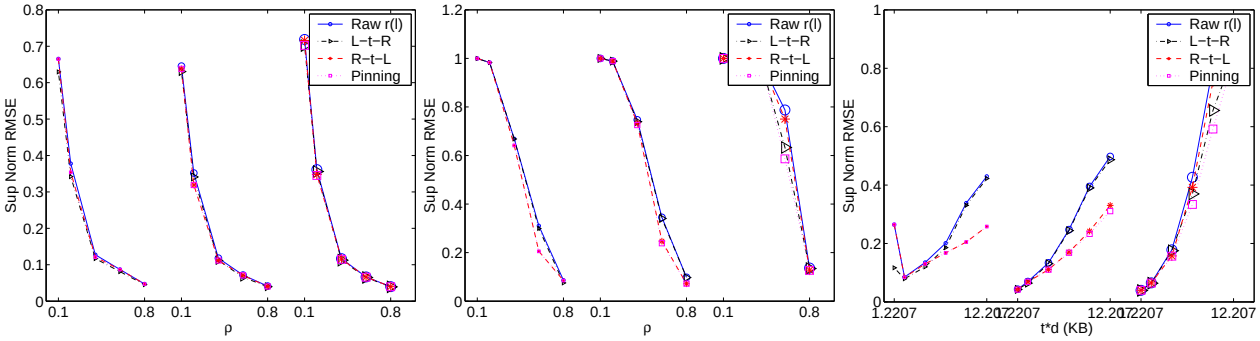

Figure 16: Sup-norm performance of various composite estimators. The ρ access is repeated for three window sizes: ($t/50$, $t/10$, $t/2$). (a) As a function of ρ, $t = 1.67 p_{max}/d$; (b) As a function of ρ, $t = 6.67 p_{max}/d$; (c) Dependence on t, with $\rho = 0.6$.

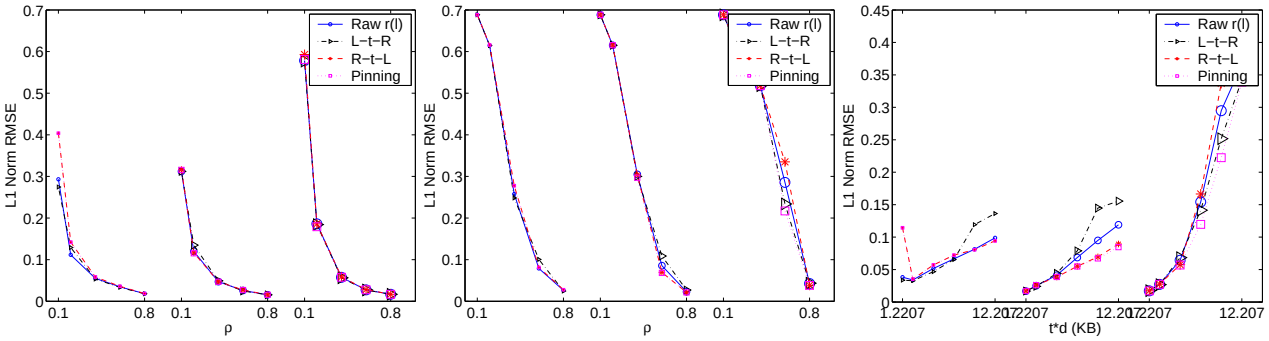

Figure 17: L1-norm performance of various composite estimators. The ρ access is repeated for three window sizes: ($t/50$, $t/10$, $t/2$). (a) As a function of ρ, $t = 1.67 p_{max}/d$; (b) As a function of ρ, $t = 6.67 p_{max}/d$; (c) Dependence on t, with $\rho = 0.6$.

which controls h and therefore C. Figure 14(a) shows performance as a function of ρ for a fixed t. We see that the naive estimator $\mathbf{r} = 0$, which blindly applies the packet pair heuristic, performs very poorly, whereas the Class 1 estimator with $\mathbf{r} = t - x$ performs as well as the sophisticated variants once ρ exceeds 0.2. This indicates that for these (ρ, t) combinations there is sufficient data, and the methods are effectively defaulting to using $\mathbf{r} = t - x$. We see a steady improvement as ρ increases, as increasing data leads to lower variance and hence MSE. Here the raw composite estimator was used.

Figure 14(b) shows the effect of increasing t by a factor of 4. The effective loss of available data sees $\mathbf{r} = t - x$ performing poorly now until ρ is at least 0.8, since there is little mass to the right of the strip, except at very high utilisation. At $\rho = 0.1$, data is so scarce that all methods have errors which are equal becaue they are the worst possible, namely equal to 1 for some l. Significant improvement is achieved by using the tuned Class 2 estimators which enter into the strip, not not too far, provided that ρ is high enough. The adaptive estimators all perform well, however the need to choose m wisely is apparent: as ρ increases, larger m performs better, although at still larger ρ, they all default to $t-x$ as so perform identically. The adaptive and constant variants of $m = 8$ perform similarly, although the adaptive one is consistently slightly better. Finally, the raw composite estimator shows uniformly good results, demonstrating a satisfying adaptivity to the amount of data available.

Figure 14(c) shows the effect of increasing t at fixed $\rho = 0.6$ (the first and fourth t values correspond to those of plots (a) and (b) respectively). Not surprisingly, all methods perform worse at greater t, as there is less effectively less data available (their are some exceptions in a few cases at the largest two t values. We do not fully understand these, but they are probably due to the definition of the 'truncated L1' measure which does not scale appropriately as $C(l)$ evolves with t. In particular, there is a need to exclude the tail not only at large l but also at small. As the left tails grows in length as t (recall that $\mathbb{E}[C]$ is proportional to t, this will reduce the measure at large enough t. There may also be some contribution due to variability in our estimation of these performance curves, which use only $N = 500$ probes). At larger t, the cross traffic is effectively less bursty as far as observations by probes are concerned. Just as in the case of low utilisation, reduced burstiness robs the probes of the variation in delays they require to obtain data from the required region. In particular it is more difficult to find mass to the right of the strip. From plot (c) we see that, although errors build quite rapidly for larger t, good estimator design has the potential to slow this growth substantially, whereas the extremes $t = 0$ and $t = t - x$ perform very badly in general. Again we find that the raw composite estimator successfully adapts to the changing traffic conditions, whereas for the adaptive estimator m must be chosen appropriately.

Figure 15 shows the analogous plots but using the L1 measure. The same general results are found, although the root MSE errors are smaller, as this is no longer the worse case error. The composite estimator performs even better with respect to this metric, and therefore seems to be the best estimator overall.

As the raw composite estimator shows considerable promise, we now subject it and its monotonized variants to a more detailed performance study. Figures 16 and 17 explore the performance over (ρ, t) space in a similar way to before. In each plot, the 3 different monotonicity algorithms, and the raw composite, are compared, for each of three different window sizes. To avoid clutter, the results for different window sizes have been displaced horizontally using duplicate ρ axes. It is important to understand that here the results pertain to the (root) MSE of a single sample. Thus, the l value at which the Sup is found will vary from sample to sample. In contrast, Figure 11 showed displayed expected results for each l fixed.

The results of Figures 16 and 17 show a remarkable lack of variation across both the methods and the window sizes. Part of this is understandable. For each window size, each monotonicity algorithm uses the same underlying raw $r(l)$, and so shares the same environment in terms of data availability. For very large t however we do see that windows sizes that are too large perform poorly, which can be understood by noting that when data poor, selecting smaller **r** is necessary to capture the few angles that are available, whereas larger w will favour the algorithm triggering at larger **r**. We also find that when there is a difference, Data Pinnning performs best among the monotonic estimators, as we might expect from the insights of Figure 11. In fact for these per-sample metrics, it performs even better than the raw composite, although the differences are typically small. This apparent contradiction can be explained by noticing that looking at ensemble performance with l fixed, or not, corresponds in fact to two very different metrics which are not obliged to correspond. It seems that, although at a fixed l raw is clearly superior (recall Figure 11), sample functions are so variable that good behaviour at some l is systematically compensated by worse behaviour elsewhere, resulting in a final performance which is less dependent on the details of the monotonicity algorithm than one might have supposed (we speculate that the different variants effectively select from the same underlying sample functions, but select a different one in different samples). This is good news in the sense that it seems that the low bias of the raw composite estimator can effectively be achieved in a monotonized version, and in particular, the results of Figure 14 and Figure 15 for the raw composite, which correspond to the central plots in Figures 16 and 17, still hold for the monotonic variants, especially Data Pinning. Finally, we note that L2R performed best in some data rich cases (small t and large ρ) not shown here.

It is worth remembering that in data poor cases, sample functions can be extremely crude and large errors are made by all methods, for example sample CDFs containing only 1 or 2 jumps, which typically have a Sup error of 1. When collecting statistics over large numbers of samples however to estimate the MSE, this becomes less apparent as the jump positions change, reinstating details of the structure of the CDF which are absent from individual samples. On the other hand, if data is plentiful, then the sample functions are relatively detailed and almost monotonic from the beginning, and so the effect of the different algorithms is not very large. Thus, the differences between algorithms and methods manifest themselves in the intermediate zone where data is neither too scarce, nor plentiful.

5 Trace Analysis

In this section we use real data from core Internet routers to demonstrate the performance of our estimators with real traffic. We use both passive trace data, and a unique data set involving simultaneous passive capture and active probing.

5.1 Trace Driven Simulations

We use the same queueing system as that used in Section 4.3. We generated the cross traffic by replaying the traces from the full router experiment described in [6], to which we had access. This experiment recorded all packets entering every interface of a router over a 13-hour time period. We model the output buffer of a particular OC-3 output interface that has a reasonably high utilization, and replay the cross traffic that passed through it.

5.1.1 Data Overview

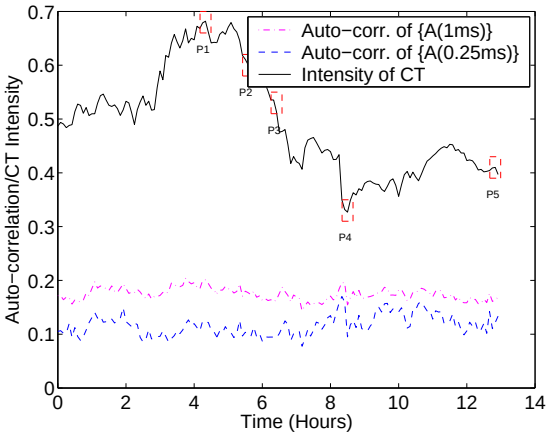

Figure 18: Traffic characteristics at the OC-3 link. Top: byte intensity measured over 1 [sec] intervals. Four 5 [min] long regions are identified with a spread of ρ values. Bottom plots: lag-1 autocorrelation estimates, based on looking at $A(t)$ for $t = 1$ms and $t = 0.25$ms, calculated over 5 minute intervals.

By using these traces, we do more than make use of a source of realistic cross traffic. Because of the complete monitoring, fine grained detail of all input packets destined to the chosen output are also available. It is therefore possible to reproduce the true (B, C) values, and compare them to those predicted by the estimator $\hat{h}(k, l)$.

As they are real traces, we did not control the ρ values in the traces a priori. To obtain a spread of values, we first observe the actual traffic intensity, averaged over 1 second intervals, of bytes to the OC-3 link. As shown in the upper plot in Figure 18, these range from $\rho = 0.4$ to 0.7. Four 5 minute portions of the trace, identified in order of increasing ρ as $P1$, $P2$, $P3$, $P5$ (note the inversion) and $P4$, were chosen to provide a spread across this range. We use the cross-traffic arrival processes corresponding to these to drive the simulations.

We make an attempt to measure the degree to which the cross traffic obeys the i.i.d. assumptions of $A(t)$ at each of $t = 0.25$ms and $t = 1$ms, by estimating the 1st lag correlation coefficient

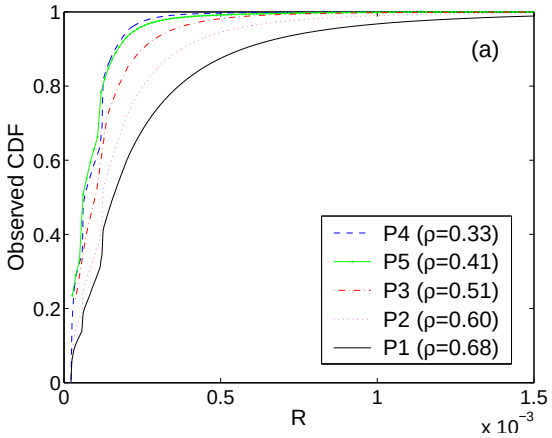

Figure 19: Marginals of R for the trace portions of Figure 18.

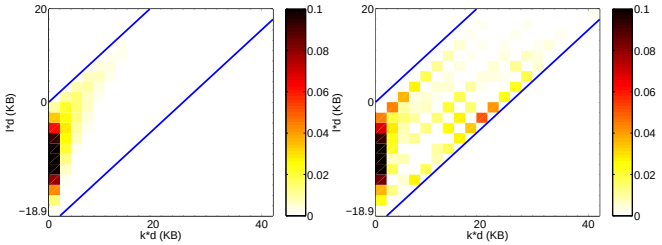

Figure 20: Comparison of measured and estimated $h(k,l)$ for $P1$ ($300,000$ probe packets), with $d = 10$ and and plotting resolution of $5d$ or 0.1ms for $t = 12.9p_{max}/d$ $(0.001s)$.

for the time series corresponding to the number of bytes in non-overlapping intervals of width t across the trace. The results, shown in the lower plots in Figure 18, are small over the trace and are reasonably *consistent* with independence. Not surprisingly, it holds a little better for the larger value of $t = 1$ms.

Figure 20 shows a comparison of estimated and measured $h(k,l)$. Despite the fact that the estimation of a 2-dimensional distribution is inherently difficult, (although the density is plotted at a resolution above that of δ) it is clear that the essence of the queueing behavior as encapsulated in the packet-pair related data (B,C) is being captured by the estimator.

Figure 19(a) plots the distribution of excess packet delays, due to the trace traffic alone, for $P1$ to $P4$. There is a considerable spread across the different sub-traces, and therefore the delays experienced by probes in the corresponding experiments will likewise differ, resulting in different available data and thereby different estimator performance. To get a feeling for the trace data, note that the quantile $r_{0.99}$ is greater than 0.25ms in all cases, and not more than 1ms except in $P4$ where the intensity is highest.

5.1.2 Performance

Examples of individual estimates are given in Figure 23 to show the effects of utilisation and the number of probes n. As we can see, each has a dramatic impact on the ability of the estimator to see the numerous atoms of the true distribution, and hence to recover the structure of the CDF. The utilisation has a largest impact, and can be thought of as controlling the overall bias, whereas increasing n improves the reproduction of the CDF structure, thereby reducing error in a given sample, or alternatively provides more data which decreases variance.

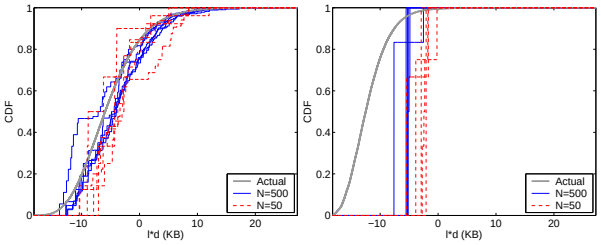

Figure 23: We plot sample estimates with different number of probes n for for utilization 68% and 31%, with $t = 12.9p_{max}/d$ $(1ms)$.

Figures 21 and 22 are the analogous results to Figures 14 and 15 using $n = 500$ and a discretization level of $d = 10$ bytes per slot. Here the t values are larger than those used before.

The relative performance across estimators is similar to that seen earlier and the absolute performance is reasonable at the values of ρ available. The t dependence explored in Figure 21(c) again reinforces the findings from the previous section. The composite estimator, using Data Pinning, is now the best performer in all plots and for both the Sup and L1 measures. However we see that for large enough t, the difference between estimators begins to diminish as they are asked to deliver the impossible, with errors which correspondingly approach the maximum of 1, first in the Sup metric, and then L1.

5.2 Active-Passive Experiment

To test our estimators in real network conditions, we conducted novel active probing experiments in which we sent probe streams along a path in a tier-1 ISP. Essentially, we performed pure 1-hop probing. In a true multi-hop path, the links other than the predominant bottleneck add noise to the observations. Other factors such as path persistence of cross-traffic also affect performance. Quantifying these is extremely path-dependent and out of our scope.

5.2.1 Experimental Setup

We chose a router in a tier-1 ISP that had an OC-192 link of utilization around 50% ($\rho = 0.5$). We injected packets across this router through an active probing device. Our active probing device was an Ixia 400T [8], a specialized hardware device that is typically used to test routers. This device sent packets on an OC-12 link that was connected via an un-utilized OC-48 link to our chosen router. Packets were addressed to suitable IP addresses so that they would be output on the chosen OC-192 link. We programmed the active probing device using a *Tcl* command-line interface provided. This allowed us to generate our probe streams. Due to factors of load, we could not generate periodic streams. Hence, we used widely-separated packet pairs with intra-pair separation t. As discussed earlier, packet pairs

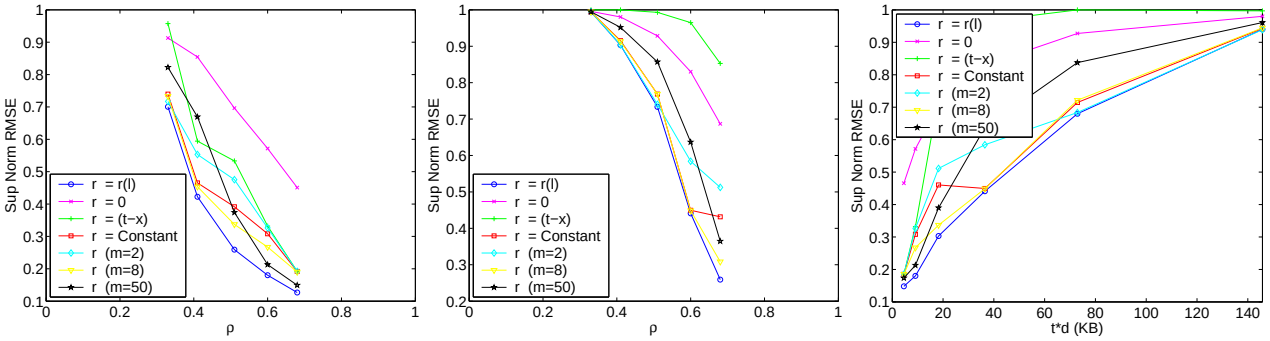

Figure 21: Estimator Sup-norm performance using router traces. We use data pinning monotonic algorithm with r(l). (a) As a function of ρ, $t = 6.45 p_{max}/d\ (500\mu s)$; (b) As a function of ρ, $t = 25.8 p_{max}/d\ (2ms)$; (c) Dependence on t, with $\rho = 0.6$.

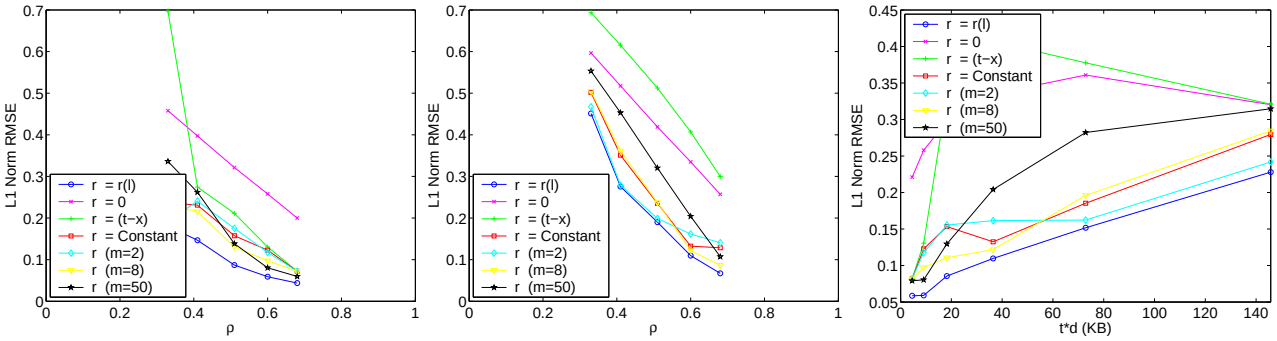

Figure 22: Estimator L1-norm performance using router traces. We use data pinning monotonic algorithm with r(l). (a) As a function of ρ, $t = 6.45 p_{max}/d\ (500\mu s)$; (b) As a function of ρ, $t = 25.8 p_{max}/d\ (2ms)$; (c) Dependence on t, with $\rho = 0.6$.

can be used as long as the consequences in Section 2.1 are true. The cross traffic is not controlled in any way.

We monitored the input OC-48 and output OC-192 links using GPS-synchronized DAG cards. This provided us with the arrival and departure times of the probe packets. The output link monitor also provided us with the departure timestamps of the intervening cross-traffic. These timestamps which were accurate to sub-microsecond levels and allowed us to calculate the empirical $A(t)$. Since the arrival timestamps of cross-traffic to the output queue could not be measured, we could not calculate $B(t)$. Hence we evaluate the performance of our estimators in calculating $C(l)$ only.

5.2.2 Performance

Due to probing load constraints we could only conduct $N = 10$ experiments at a particular time. We sent $n = 250$ packet pairs, for a range of separations $t * d$ varying from 625 bytes ($t = 500ns$) to 40KB ($32\mu s$). To achieve such small separation using an OC-12 link, we used small probes of size 40 bytes. Utilization levels could not be controlled in the experiments, however the performance of the estimators could be tested under operational conditions for a range of timescales t.

The resulting t dependence performance curves are plotted in Figure 24. Although here we control neither the queueing model nor the cross traffic, the results, at least for $td \geq 5$, are reminiscent of those seen earlier. In particular, the composite estimator performs the best, with an error at least twice that of the commonly used $\mathbf{r} = 0$ estimator over a wide range of t. Also, as t increases, the performance worsens much much more slowly using the composite or $m = 8$ adaptive estimator than the Class 1 estimator with $\mathbf{r} = t - x$. The graphs show that estimates with MSE not exceeding 0.2 can be obtained even for $t * d$ about 10 times the transmission time of 1500 bytes.

For very small t values near $1\mu s$, we see a clear *increase* in MSE. There are several possible reasons for this. First, we observed that errors in our probe generation times (essentially, our control of t) could be as large as 50% over these scales. Second, the independent increment assumption is likely to break down at these time scales.

6 Discussion and Future Work

This paper tackles the question of the in-principle potential of probing methods, viewed as a question in system identifiability. We show that, even in the 1-hop case, invertibility may be possible or not, depending on the nature of the cross traffic and the probing stream.

Using the insights gained, we defined several estimators of cross traffic, explained their principles of operation, their strengths and weaknesses, and investigated their statistical performance as a function of key parameters, such as utilisation, and the probe separation. We found that in many circumstances they performed quite well. We gained considerable insight into exactly why their performance varies as a function of cross traffic

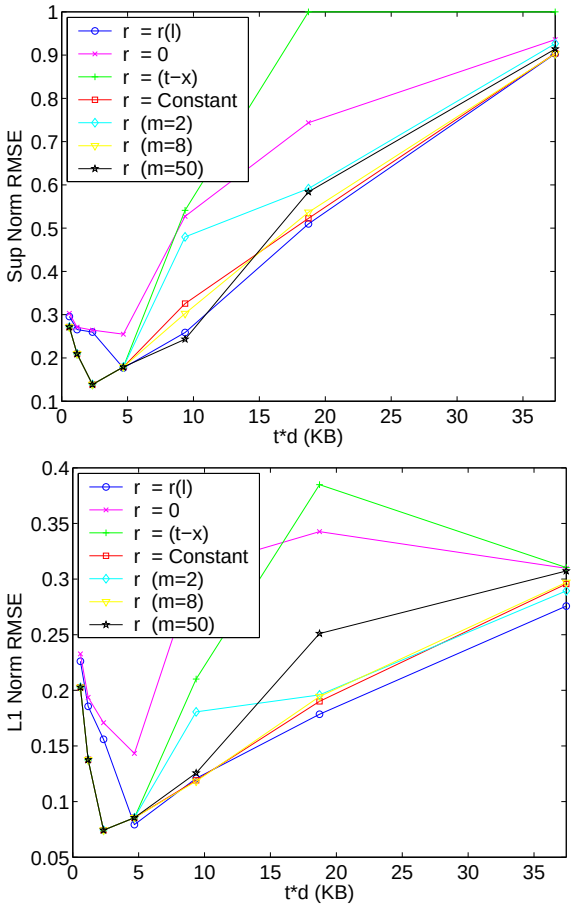

Figure 24: Estimator performance in live active/passive experiments. Dependence on t ($\rho = 0.50$).

and probing parameters. For example, the idea of 'coverage' of the cross traffic distribution by the available data can predict what kind of performance is achievable. When coverage is close to non-existant, then no method can perform well, whereas when coverage is very good, even simple methods perform well. Inbetween these extremes lie the cases where sophisticated estimation can make a very significant difference.

We then tested the estimators using real traffic of two kinds: data from highly detailed router monitoring experiment where full details of the queueing could be compared with the model predictions, and live experiments where simultaneous high precison passive monitoring again allows comparisons to be made. We found that the result using real data where comparable with those in our simulations, suggesting that the methods are robust enough to be useful in practice.

Our work can be seen as a generalisation of packet pair, where useful information can be extracted for far greater probe separations than has been supposed in the past.

Active probing for capacity estimation of the bottleneck link bandwidth has been studied in [4, 9, 12, 16]. One of these techniques can be used to determine μ in our scheme, if necessary. Available bandwidth estimation has been the focus of many prior works [7, 11, 21]. Essentially, they all focus on estimating the first-order moment of the cross-traffic arrival process. In contrast, our goal is to estimate the entire cross-traffic arrival process. Also, many of these [7, 11] assume fluid models for cross-traffic unlike our assumptions of a discrete packet-based system. Recent work [13, 14] analyzing single-hop available bandwidth techniques has characterized the discrepancy between fluid flow and packet-based models. They do not focus on estimating cross-traffic properties. Another class of network inference work is the network tomography literature (see [22] and references therein) which is mostly focused on estimating delays at queues.

Some prior work [1, 18, 20] has focused on cross-traffic estimation. To our knowledge, all of them assume a specific cross-traffic arrival process, e.g. multi-fractal wavelet, Poisson. Based on this assumption, they attempt to estimate the parameters describing these processes. Our work is more general in two ways. First, our assumptions are more general and encompass a variety of processes. Our results with real experiments also bears this out. Second, our system framework makes it possible to be adapted to other kinds of assumptions too.

There are many directions for future work. These include formalising the performance of the estimators presented here, and further refining the estimators themselves. The question of probe stream design enters naturally at this point and will also be the focus of future work. Finally, the validation of the stationary assumptions assumption in practice, and robustness to the breaking of it, is an important point to investigate.

References

[1] S. Alouf, P. Nain, and D. Towsley. Inferring Network Characteristics via Moment-based Estimators. In *Proc. of IEEE Infocom*, 2001.

[2] F. Baccelli and P. Bremaud. *Elements of Queueing Theory*. Springer Verlag, Applications of Mathematics, second edition, 2003.

[3] C. Dovrolis, P. Ramanathan, and D. Moore. What do packet dispersion techniques measure? In *Proceedings of IEEE Infocom'01*, Anchorage, Alaska, April 22–26 2001.

[4] C. Dovrolis, P. Ramanathan, and D. Moore. What do Packet Dispersion Techniques Measure? In *Proc. of INFOCOM*, 2001.

[5] N. Hohn, D. Veitch, K. Papagiannaki, and C. Diot. Bridging router performance and queuing theory. In *Proceeding of ACM Sigmetrics 2004 Conference on the Measurement and Modeling of Computer Systems*, New York, 12–16 June 2004.

[6] N. Hohn, D. Veitch, K. Papagiannaki, and C. Diot. Bridging Router Performance and Queuing Theory. In *Proc. of ACM SIGMETRICS*, June 2004.

[7] N. Hu and P. Steenkiste. Evaluation and Characterization of Available Bandwidth Probing Techniques. *IEEE Journal on Selected Areas in Communications, Special Issue on Internet and WWW/Measurement, Mapping, and Modeling*, August 2003.

[8] Ixia. `http://www.ixiacom.com`.

[9] V. Jacobson. *Pathchar – a tool to infer characteristics of Internet paths*, available at: `http://www.employees.org/bmah/Software/pchar/` edi-

tion, 1997.

[10] M. Jain and C. Dovrolis. End-to-End Available Bandwidth: Measurement Methodology, Dynamics, and Relation with TCP Throughput. In *Proceedings of ACM SIGCOMM'02*, Pittsburgh, Pennsylvania, Aug 19-23 2002.

[11] M. Jain and C. Dovrolis. End-to-end Available Bandwidth: Measurement Methodology, Dynamics, and Relation with TCP Throughput. *IEEE/ACM Transactions on Networking*, 11(4):537–549, 2003.

[12] R. Kapoor, K.-J. Chen, L. Lao, M. Gerla, and M. Y. Sanadidi. CapProbe: A Simple and Accurate Capacity Estimation Technique. In *Proceedings of ACM SIGCOMM*, 2004.

[13] X. Liu, K. Ravindran, B. Liu, and D. Loguinov. Single-Hop Probing Asymptotics in Available Bandwidth Estimation: Sample-Path Analysis. In *Proc. of ACM SIGCOMM IMC*, October 2004.

[14] X. Liu, K. Ravindran, and D. Loguinov. What Signals do Packet Pair Dispersions Carry? In *Proc. of IEEE Infocom*, 2005.

[15] A. Pásztor. *Accurate Active Measurement in the Internet and its Applications*. PhD thesis, University of Melbourne, Victoria 3010, Australia, 2003.

[16] A. Pásztor and D. Veitch. Active Probing using Packet Quartets. In *Proc. ACM SIGCOMM Internet Measurement Workshop (IMW-2002)*, pages 293–305, Marseille, Nov 6–8 2002.

[17] A. Pásztor and D. Veitch. On the scope of end-to-end probing methods. *IEEE Communications Letters*, 6(11):509–511, Nov. 2002.

[18] V. Ribeiro, M. Coates, R. Riedi, S. Sarvotham, and R. G. Baraniuk. Multifractal cross-traffic estimation. In *Proceedings of the ITC Specialist Seminar on IP Traffic Measurement, Modelling and Management 2000*, pages 15(1–10), Monterey, CA, 2000.

[19] V. Ribeiro, R. Riedi, J. Navratil, and L. Cottrell. pathChirp: Efficient Available Bandwidth Estimation for Network Paths. In *PAM 2003, Passive and Active Measurement Workshop*, La Jolla, California, April 6–8 2003.

[20] V. Sharma and R. Mazumdar. Estimating traffic parameters in queueing systems with local information. *Performance evaluation*, 32:217–230, 1998.

[21] J. Strauss, D. Katabi, and F. Kaashoek. A Measurement Study of Available Bandwidth Estimation Tools. In *Proceedings of the 2003 ACM SIGCOMM Conference on Internet Measurement*, pages 39–44, 2003.

[22] Y. Tsang, M. Yildiz, R. Nowak, and P. Barford. Network Radar: Tomography from Round Trip Time Measurement. In *ACM Internet Measurement Conference*, pages 175–180, Taormina, Sicily, Italy, Oct 2004.

A Ergodicity

In Model 1, the independence between A and T_n and the strong Markov property imply that the sequence $\{(B_n, C_n)\}$ is i.i.d. and the law of C has an infinite support (thanks to the assumption that A has an infinite support). Hence $\{R_n\}$ is an irreducible and ergodic Markov chain on the positive half line provided the rate condition

$$\frac{x}{E(T_{n+1} - T_n)} + E(A(0,1]) < 1$$

is satisfied.

In Model 2, the sequence $\{(B_n, C_n)\}$ is i.i.d. by assumption, so that $\{R_n\}$ is a Markov chain. For this Markov chain to be irreducible and ergodic on the whole half line, under the rate condition given above, it is enough that $P(A[0,t) = 0) > 0$ and that $P(A[0,t) > t - x) > 0$.

B Meaning of B, example from Network Calculus

Assume that A is (σ, ρ)-regulated (in the sense of network calculus [2], with $\rho \leq 1$, that is that it obeys

$$A([s,t)) \leq (t-s)\rho + \sigma, \quad \forall s \leq t,$$

where σ is the burst parameter and ρ the rate parameter. Then

$$\begin{aligned} B_n &= \sup_{s \in [T_n, T_{n+1}]} A([s, T_{n+1})) - (T_{n+1} - s) \\ &\leq \sup_{s \in [T_n, T_{n+1}]} (\sigma + (T_{n+1} - s)\rho - (T_{n+1} - s)) \\ &= \sigma. \end{aligned}$$

where the last inequality may be reached within the class of all (σ, ρ)-regulated measures.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifi que
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399